

Cell



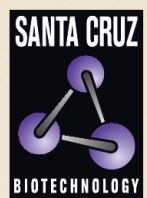
Volume 159
Number 5

November 20, 2014

www.cell.com

**Cradle for Endosome
Biogenesis
A Century of
X-Ray Structures**

SANTA CRUZ BIOTECHNOLOGY, INC.



The Power to Question

INTRODUCING

CRISPR/Cas9

KNOCKOUT PLASMIDS

FAST, EFFICIENT GENOME
EDITING TOOLS

EXCLUSIVELY ONLINE AT:

SCBT.COM



Smooth Operator

The new Eppendorf micromanipulators TransferMan® 4r and InjectMan® 4

The new Eppendorf micromanipulators have been developed to simplify your work and increase the quality of your micromanipulation results. Based on their ergonomic design, the robust and reliable devices provide excellent performance for a wide range of applications.

- > Unprecedented movement control and maximum stability
- > Pre-defined functions for various applications to optimize your workflow
- > Connection with new Eppendorf electronic microinjectors FemtoJet® 4i/4x and Eppendorf PiezoXpert®

www.eppendorf.com/cellmanipulation



Abstract
submission
deadline:

August 22
2014

Early bird
deadline:

October 3
2014

Stem Cell Energetics will bring together academic researchers, clinicians and drug developers working in the fields of metabolism and stem cells to provide a forum for interdisciplinary discussion and cross-collaboration amongst experts in these rapidly advancing fields.

Symposium themes:

- Metabolic regulation of stem cell self renewal
- Niche influence on stem cell metabolism
- Energetics of stem cell flux
- Mitochondria regulation of stem cell fate
- Metabolites in stem cell epigenetics and reprogramming
- Altered stem cell metabolism in aging and disease

Organizers

Nikla Emambokus, Editor, *Cell Metabolism*

Sheila Chari, Scientific Editor, *Cell Stem Cell*

Emmanuelle Passegue, University of California, San Francisco, USA

Michael Teitell, University of California, Los Angeles, USA

www.cell-symposia-stem-cell-energetics.com

Cell

Editor

Emilie Marcus

Senior Deputy Editor

Elena Porro

Deputy Editors

Karen Carniol
Robert Kruger

Scientific Editors

Mirna Kvajo
Cindy C. Lu
Steve Mao
João Monteiro
Sri Devi Narasimhan
Lara Szewczak
Jiaying Tan

Senior Managing Editor

Andy Smith

Managing Editor

Trina Arpin

Deputy Managing Editor

Julie Fiorilla

Art Program Manager

Andrew A. Tang

**Managing Designer
and Illustrator**

Yvonne Blanco

Illustrator

Kip B. Lyall

Editorial Assistant

Anna Hofvander

Editorial Board

C. David Allis
Geneviève Almouzni
Angelika Amon
Silvia Arber
Johan Auwerx
Richard Axel
Cori Bargmann
Konrad Basler
David Baulcombe
Yasmine Belkaid
Jeffrey Benovic
Wendy Bickmore
Elizabeth Blackburn
Joan Brugge
Lewis Cantley
Xuetao Cao
Joanne Chory
David Clapham
Andrew Clark
Hans Clevers
Pascale Cossart
Benjamin Cravatt
George Daley
Jeff Dangl
Ted Dawson
Karl Deisseroth
Hongkui Deng
Pier Paolo di Fiore
Ivan Dikic
Julian Downward
Bruce Edgar
Steve Elledge
Anne Ephrussi
Ronald Evans
Witold Filipowicz
Marco Foiani
Elaine Fuchs
Daniel Geschwind

Yukiko Goda
Stephen Goff
Joe Goldstein
Douglas Green
Leonard Guarente
Taekjip Ha
Daniel Haber
Ulrike Heberlein
Mark Hochstrasser
Erika Holzbaur
Huck Hui Ng
Tony Hunter
James Hurley
Richard Hynes
Trey Ideker
Thomas Jessell
Tarun Kapoor
Narry Kim
Mary-Claire King
David Kingsley
Frank Kirchhoff
Robert Lamb
Mark Lemmon
Beth Levine
Wendell Lim
Jennifer Lippincott-Schwartz
Dan Littman
Richard Losick
Scott Lowe
Tom Maniatis
Matthias Mann
Kelsey Martin
Joan Massagué
Satyajit Mayor
Ruslan Medzhitov
Craig Mello
Tom Misteli
Tim Mitchison
Danesh Moazed

Alex Mogilner
Paul Nurse
Roy Parker
Dana Pe'er
Kathrin Plath
Carol Prives
Klaus Rajewsky
Lalita Ramakrishnan
Venki Ramakrishnan
Rama Ranganathan
Anne Ridley
Alexander Rudensky
Helen Saibil
Charles Sawyers
Ueli Schibler
Joseph Schlessinger
Hans Schöler
Trina Schroer
Kevan Shokat
Nahum Sonenberg
James Spudich
Deepak Srivastava
Azim Surani
Keiji Tanaka
Craig Thompson
Robert Tjian
Ulrich von Andrian
Gerhard Wagner
Jonathan Weissman
Matthew Welch
Tian Xu
Shinya Yamanaka
Richard Young
Marino Zerial
Xiaowei Zhuang
Juleen Zierath
Huda Zoghbi

Cell Office

Cell, Cell Press, 600 Technology Square, 5th Floor, Cambridge, Massachusetts 02139
Phone: (+1) 617 661 7057, Fax: (+1) 617 661 7061, E-mail: celleditor@cell.com
Online Publication: <http://www.cell.com>

Cell (ISSN 0092-8674) is published biweekly by Cell Press, 600 Technology Square, 5th Floor, Cambridge, Massachusetts 02139. The institutional subscription rate for 2014 is \$1,682 (US and Canada) or \$1,895 (elsewhere). The individual subscription rate is \$246 (US and Canada) or \$371 (elsewhere). The individual copy price is \$50. Periodicals postage paid at Boston, Massachusetts and additional mailing offices. Postmaster: send address changes to Elsevier Customer Service Americas, Cell Press Journals, 3251 Riverport Lane, Maryland Heights, MO 63043, USA.

Ⓒ The paper used in this publication meets the requirements of ANSI/NISO Z39.48-1992 (Permanence of Paper). Printed by Dartmouth Printing Company, Hanover, NH.



Abstract
submission
deadline:

August 22
2014

Early bird
deadline:

October 10
2014

Join us in Belgium in December 2014 as we assemble some of the leading investigators examining various aspects of type 2 immunity in order to identify commonalities as well as contrasts of the cellular and molecular participants. This cross-fertilization will facilitate insights into this group of biological responses in health and disease.

Symposium themes:

- Role of innate immune cells in type 2 immunity
- Coordinating innate and adaptive type 2 inflammation
- Type 2 immunity in asthma and allergy and airway infection
- Type 2 immunity in microbial infection
- Role of commensal flora in influencing type 2 immunity

Keynote Speaker

Alan Sher, *NIAID, USA*

Visit our website to check out the full speaker line up

Organizers

Bruce Koppelman, Scientific Editor, *Immunity*

Ella Hinson, Scientific Editor, *Cell Host & Microbe*

Bart Lambrecht, *VIB and Ghent University, Belgium*

David Artis, *University of Pennsylvania, USA*

www.cell-symposia-type2-immunity.com

www.cell.com/symposia

Cell Press

CEO

Emilie Marcus

Vice President of Business Development

Joanne Tracy

Vice President of Operations

Keith Wollman

Publishing Directors

Peter Lee

Deborah Sweet

Editorial Director, Reviews Strategy

Katja Brose

Editorial Director, Content Development

Elena Porro

Director of Production

Meredith Adinolfi

Director of Marketing

Jonathan Atkinson

Media Relations Manager

Mary Beth O'Leary

Display Advertising

For advertising, sponsorship, and reprint opportunities, contact Jonathan Christison (jchristison@cell.com) or visit <http://advertisers.cell.com/contact-us>.

Recruitment Advertising

For recruitment advertising, visit <http://recruitmentads.cell.com/contact-us>.

©2014 Elsevier Inc. All rights reserved.

This journal and the individual contributions contained in it are protected under copyright by Elsevier Inc., and the following terms and conditions apply to their use:

Photocopying

Single photocopies of single articles may be made for personal use as allowed by national copyright laws. Permission of the Publisher and payment of a fee is required for all other photocopying, including multiple or systematic copying, copying for advertising or promotional purposes, resale, and all forms of document delivery. Special rates are available for educational institutions that wish to make photocopies for nonprofit educational classroom use.

Permissions

For information on how to seek permission, visit <http://www.elsevier.com/permissions> or call +44 1865 843830 (UK) / +1 215 239 3804 (US).

Derivative Works

Subscribers may reproduce tables of contents or prepare lists of articles including summaries for internal circulation within their institutions. Permission of the Publisher is required for resale or distribution outside the institution. Permission of the Publisher is required for all other derivative works, including compilations and translations (please consult <http://www.elsevier.com/permissions>).

Electronic Storage or Usage

Permission of the Publisher is required to store or use electronically any material contained in this journal, including any article or part of an article (please consult www.elsevier.com/permissions). Except as outlined above, no part of this publication may be reproduced, stored in a retrieval system, or transmitted in any form or by any means, electronic, mechanical, photocopying, recording, or otherwise, without prior written permission of the Publisher.

Notice

No responsibility is assumed by the Publisher for any injury and/or damage to persons or property as a matter of products liability, negligence, or otherwise, or from any use or operation of any methods, products, instructions, or ideas contained in the material herein. Because of rapid advances in the medical sciences, in particular, independent verification of diagnoses and drug dosages should be made. Although all advertising material is expected to conform to ethical (medical) standards, inclusion in this publication does not constitute a guarantee or endorsement of the quality or value of such product or of the claims made of it by its manufacturer.

Reprints

Article reprints are available through Elsevier's reprint service; for information, contact Derrick Imasa (e-mail: d.imasa@elsevier.com; ph: +1 212 633 3874).

Subscription Orders and Inquiries

Mail, fax, or e-mail address changes to Elsevier Customer Service Americas, allowing 4–6 weeks for processing. Lost or damaged issues will be replaced, subject to availability, if Cell Press is notified within the claim period (US and airmail delivery: 3 months from issue date; surface delivery: 4 months from issue date). Periodical delivery in the US can take up to 3 weeks. Airmail delivery can take 2–4 weeks. Mailing address: Elsevier Customer Service Americas, Cell Press Journals, 3251 Riverport Lane, Maryland Heights, MO 63043, USA. Toll-free phone within USA/Canada: +1 866 314 2355; phone for outside US/Canada: +1 314 447 8880; fax: +1 314 447 8029; e-mail: subs@cell.com.

Funding Body Agreements and Policies

Elsevier has established agreements and developed policies to allow authors whose articles appear in journals published by Elsevier to comply with potential manuscript archiving requirements as specified as conditions of their grant awards. To learn more about existing agreements and policies, please visit <http://www.cell.com/cellpress/FundingBodyAgreements>.



It only took us

125 years
to write this book...

www.cellsignal.com/cell125

CST Guide: Pathways and Protocols. From the inception of the antibody as a research tool in the 1890s, to the most up-to-date research, applications, and tools, this is your complete resource for cellular research.

For Research Use Only. Not For Use In Diagnostic Procedures.

© 2014 Cell Signaling Technology, Inc. Cell Signaling Technology® and CST™ are trademarks of Cell Signaling Technology, Inc.



Cell Signaling
TECHNOLOGY®

Leading Edge

Cell Volume 159 Number 5, November 20, 2014

IN THIS ISSUE

SELECT

961 History Lessons from Bacteria

EDITORIAL

965 Credibility and Reproducibility *E. Marcus*

VOICES

967 What's Your Favorite Crystal Structure?

COMMENTARY

969 Accelerating HIV-1 Vaccine Efficacy Trials *D.H. Barouch and N.L. Michael*

PREVIEWS

973 Genetic Variation Meets
Replication Origins *R.J. Hause and J. Shendure*

975 HIV-1 Packing to Leave *B.R. Cullen*

977 A Molecular Gauge for Nitrogen
Economy in Plants *T.R. Fallon and J.-K. Weng*

PERSPECTIVE

979 The Ribosome Emerges from a Black Box *V. Ramakrishnan*

REVIEWS

985 A Tale of Chromatin and Transcription
in 100 Structures *P. Cramer*

995 A Glimpse of Structural Biology
through X-Ray Crystallography *Y. Shi*

TIMELINE

1230 Cell Line: 2004–2014

(continued)

PROTECT

Detect, eliminate and prevent cell culture contamination



Before mycoplasma contamination compromises your cell cultures, stop it. Biological Industries has everything you need.

Disinfect incubators, sterile benches and other surfaces regularly with our pharmacidal spray.

- Prevents contamination and growth of fungi, bacteria, mycoplasma and viruses
- Non-toxic, biodegradable and non-corrosive

Keep water baths clean and clear.

- **Aquaguard-1:** for disinfecting water baths of CO₂ incubators
- **Aquaguard-2:** for disinfecting various kinds of water baths from bacteria and fungi

Detect suspected contamination with our EZ-PCR Mycoplasma Test Kit.

- Complete, ready-to-use—including TAQ polymerase and positive template control cultures
- High sensitivity and selectivity—selective primers for over 20 species of mycoplasma DNA
- Results in less than 5 hours

Treat mycoplasma contamination effectively when it happens.

- **BIOMYC-1:** based on the antibiotic tiamutin
- **BIOMYC-2:** based on the antibiotic minocycline
- **BIOMYC-3:** based on the antibiotic ciprofloxacin

Protect your cell cultures 24/7 with mycoplasma detection and treatment products from Biological Industries. Available in the U.S. from Rainbow Scientific, Inc.



Download brochure



Visit www.shoprainbowsscientific.com for easy online ordering or call 860.298.8382.



BI
Biological Industries
Culture of Excellence
www.bioind.com

Leading North American supplier of state-of-the-art products

Articles

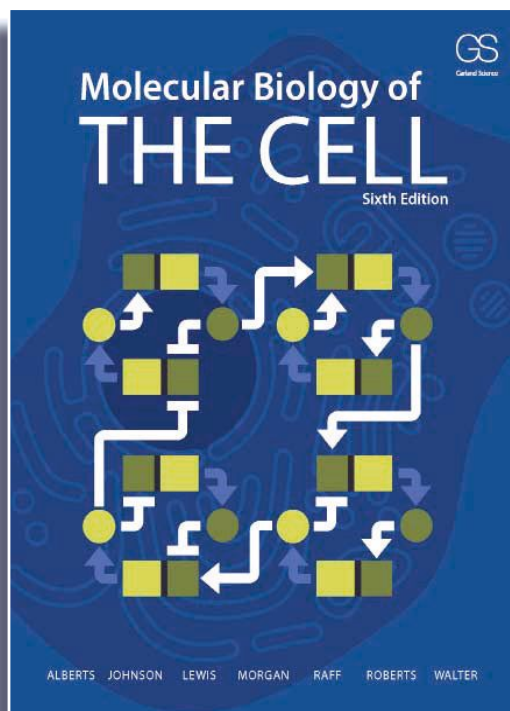
Cell Volume 159 Number 5, November 20, 2014

- 1015 Genetic Variation in Human DNA Replication Timing
A. Koren, R.E. Handsaker, N. Kamitaki, R. Karlić, S. Ghosh, P. Polak, K. Eggan, and S.A. McCarroll
- 1027 ER Contact Sites Define the Position and Timing of Endosome Fission
A.A. Rowland, P.J. Chitwood, M.J. Phillips, and G.K. Voeltz
- 1042 The Structural Basis of Substrate Recognition by the Eukaryotic Chaperonin TRiC/CCT
L.A. Joachimiak, T. Walzthoeni, C.W. Liu, R. Aebersold, and J. Frydman
- 1056 A Memory System of Negative Polarity Cues Prevents Replicative Aging
F. Meitinger, A. Khmelinskii, S. Morlot, B. Kurtulmus, S. Palani, A. Andres-Pons, B. Hub, M. Knop, G. Charvin, and G. Pereira
- 1070 Proinflammatory Signaling Regulates Hematopoietic Stem Cell Emergence
R. Espín-Palazón, D.L. Stachura, C.A. Campbell, D. García-Moreno, N. Del Cid, A.D. Kim, S. Candel, J. Meseguer, V. Mulero, and D. Traver
- 1086 RACK1 Controls IRES-Mediated Translation of Viruses
K. Majzoub, M.L. Hafirassou, C. Meignin, A. Goto, S. Marzi, A. Fedorova, Y. Verdier, J. Vinh, J.A. Hoffmann, F. Martin, T.F. Baumert, C. Schuster, and J.-L. Imler
- 1096 Global Changes in the RNA Binding Specificity of HIV-1 Gag Regulate Virion Genesis
S.B. Kutluay, T. Zang, D. Blanco-Melo, C. Powell, D. Jannain, M. Errando, and P.D. Bieniasz
- 1110 lncRNA Directs Cooperative Epigenetic Regulation Downstream of Chemokine Signals
Z. Xing, A. Lin, C. Li, K. Liang, S. Wang, Y. Liu, P.K. Park, L. Qin, Y. Wei, D.H. Hawke, M.-C. Hung, C. Lin, and L. Yang
- 1126 CDK7 Inhibition Suppresses Super-Enhancer-Linked Oncogenic Transcription in MYCN-Driven Cancer
E. Chipumuro, E. Marco, C.L. Christensen, N. Kwiatkowski, T. Zhang, C.M. Hatheway, B.J. Abraham, B. Sharma, C. Yeung, A. Altabef, A. Perez-Atayde, K.-K. Wong, G.-C. Yuan, N.S. Gray, R.A. Young, and R.E. George
- 1140 Circadian Enhancers Coordinate Multiple Phases of Rhythmic Gene Transcription In Vivo
B. Fang, L.J. Everett, J. Jager, E. Briggs, S.M. Armour, D. Feng, A. Roy, Z. Gerhart-Hines, Z. Sun, and M.A. Lazar
- 1153 A Variety of Dicer Substrates in Human and *C. elegans*
A. Rybak-Wolf, M. Jens, Y. Murakawa, M. Herzog, M. Landthaler, and N. Rajewsky
- 1168 Unraveling the Biology of a Fungal Meningitis Pathogen Using Chemical Genetics
J.C.S. Brown, J. Nelson, B. VanderSluis, R. Deshpande, A. Butts, S. Kagan, I. Polacheck, D.J. Krysan, C.L. Myers, and H.D. Madhani
- 1188 A Widespread Glutamine-Sensing Mechanism in the Plant Kingdom
V.-R. Chellamuthu, E. Ermilova, T. Lapina, J. Lüddecke, E. Minaeva, C. Herrmann, M.D. Hartmann, and K. Forchhammer

(continued)

NEW SIXTH EDITION

MOLECULAR BIOLOGY OF THE CELL



BRUCE ALBERTS, University of California, San Francisco, USA
ALEXANDER JOHNSON, University of California, San Francisco, USA
JULIAN LEWIS, Formerly of Cancer Research, UK
DAVID MORGAN, University of California, San Francisco, USA
MARTIN RAFF, University College London, UK
KEITH ROBERTS, Emeritus, University of East Anglia, UK
PETER WALTER, University of California, San Francisco, USA

As the amount of information in biology expands dramatically, it becomes increasingly important for textbooks to distill the vast amount of scientific knowledge into concise principles and enduring concepts. As with previous editions, *Molecular Biology of the Cell, Sixth Edition* accomplishes this goal with clear writing and beautiful illustrations. The Sixth Edition has been extensively revised and updated with the latest research in the field of cell biology, and it provides an exceptional framework for teaching and learning.

December 2014

1,464 pages • 1,492 illustrations

Hardback • 978-0-8153-4432-2 • \$169

Loose-leaf • 978-0-8153-4524-4 • \$125

E-books available, including rentals

For more information and to purchase at a **30% DISCOUNT**, visit www.garlandscience.com/mboc6 and apply **PROMO CODE AGL91** at checkout.

Or visit us at **BOOTH #909** at the ASCB meeting.

CONTENTS

INTRODUCTION TO THE CELL

1. Cells and Genomes
2. Cell Chemistry and Bioenergetics
3. Proteins

BASIC GENETIC MECHANISMS

4. DNA, Chromosomes, and Genomes
5. DNA Replication, Repair, and Recombination
6. How Cells Read the Genome: From DNA to Protein
7. Control of Gene Expression

WAYS OF WORKING WITH CELLS

8. Analyzing Cells, Molecules, and Systems
9. Visualizing Cells

INTERNAL ORGANIZATION OF THE CELL

10. Membrane Structure

11. Membrane Transport of Small Molecules and the Electrical Properties of Membranes
12. Intracellular Compartments and Protein Sorting
13. Intracellular Membrane Traffic
14. Energy Conversion: Mitochondria and Chloroplasts
15. Cell Signaling
16. The Cytoskeleton
17. The Cell Cycle
18. Cell Death

CELLS IN THEIR SOCIAL CONTEXT

19. Cell Junctions and the Extracellular Matrix
20. Cancer
21. Development of Multicellular Organisms
22. Stem Cells and Tissue Renewal
23. Pathogens and Infection
24. The Innate and Adaptive Immune Systems

THEORY

- 1200 An Integrated Approach Reveals Regulatory Controls on Bacterial Translation Elongation

A.R. Subramaniam, B.M. Zid, and E.K. O'Shea

RESOURCE

- 1212 A Proteome-Scale Map of the Human Interactome Network

T. Rolland, M. Taşan, B. Charloteaux, S.J. Pevzner, Q. Zhong, N. Sahni, S. Yi, I. Lemmens, C. Fontanillo, R. Mosca, A. Kamburov, S.D. Ghiassian, X. Yang, L. Ghamsari, D. Balcha, B.E. Begg, P. Braun, M. Brehme, M.P. Broly, A.-R. Carvunis, D. Convery-Zupan, R. Corominas, J. Coulombe-Huntington, E. Dann, M. Dreze, A. Dricot, C. Fan, E. Franzosa, F. Gebreab, B.J. Gutierrez, M.F. Hardy, M. Jin, S. Kang, R. Kiros, G.N. Lin, K. Luck, A. MacWilliams, J. Menche, R.R. Murray, A. Palagi, M.M. Poulin, X. Rambout, J. Rasla, P. Reichert, V. Romero, E. Ruysinck, J.M. Sahalie, A. Scholz, A.A. Shah, A. Sharma, Y. Shen, K. Spirohn, S. Tam, A.O. Tejada, S.A. Trigg, J.-C. Twizere, K. Vega, J. Walsh, M.E. Cusick, Y. Xia, A.-L. Barabási, L.M. Iakoucheva, P. Aloy, J. De Las Rivas, J. Tavernier, M.A. Calderwood, D.E. Hill, T. Hao, F.P. Roth, and M. Vidal

ERRATA

- 1227 Molecular Architecture of the 40S•eIF1•eIF3 Translation Initiation Complex

J.P. Erzberger, F. Stengel, R. Pellarin, S. Zhang, T. Schaefer, C.H.S. Aylett, P. Cimermančić, D. Boehringer, A. Sali, R. Aebersold, and N. Ban

- 1228 ATRX Directs Binding of PRC2 to Xist RNA and Polycomb Targets

K. Sarma, C. Cifuentes-Rojas, A. Ergun, A. del Rosario, Y. Jeon, F. White, R. Sadreyev, and J.T. Lee

ANNOUNCEMENTS

POSITIONS AVAILABLE

On the cover: The endoplasmic reticulum (ER) is the largest organelle in the cell and plays numerous roles in protein synthesis and folding, stress management, calcium storage, and organelle biogenesis. New findings from Rowland et al. (1027–1041) implicate the ER as a surprising player in endosome fission, which could have important consequences for cargo sorting and trafficking. The cover shows ER tubules contacting endosomes, leading to their division. Cover image by Patrick Chitwood, Ashley Rowland, and Melissa Phillips at the University of Colorado, Boulder.



Want to learn how to prepare, submit and publish an article in a Cell Press journal?

Watch the Cell Press publication guide.



Chapter 1: Before manuscript submission



Chapter 2: After initial submission



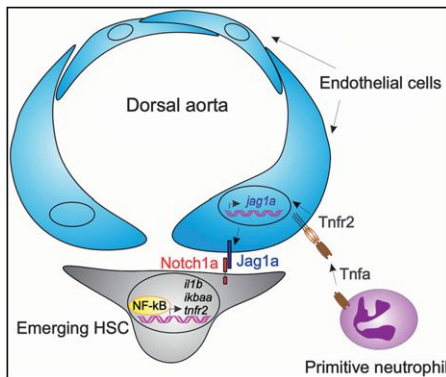
Chapter 3: Decision process



Chapter 4: After manuscript acceptance

for more information visit
www.cell.com/publicationguide





HSCs Arise in Flames

PAGE 1070

Hematopoietic stem cells (HSCs) arise from the transdifferentiation of aortic endothelium during a brief period of embryonic development. Espin-Palazon et al. show that proinflammatory signaling in the absence of infection by primitive neutrophils to the aortic endothelium regulates HSC emergence.

Whose Origins Fire First?

PAGE 1015

Rates of mutation during DNA replication are correlated with the timing of replication origin firing. Koren et al. discover that replication timing varies among humans and identify genetic variants that are *cis*-acting modifiers of replication timing, gene expression levels, and local mutation rates, thus revealing an unexpected source of variation between individuals.

Endosome DividER

PAGE 1027

Endocytic cargo and Rab GTPases are segregated to distinct domains of an endosome that maintain its identity until undergoing fission to traffic cargo. Rowland et al. now reveal that ER tubules define the position and timing of endosome fission.

Substrate Cen-TRiC Insights

PAGE 1042

Using a combination of computational and experimental approaches, Joachimiak et al. identify the substrate-binding site as well as the molecular basis of substrate recognition of the hetero-oligomeric eukaryotic chaperonin TRiC. Each TRiC subunit is found to bear a distinct, evolutionarily conserved pattern of polar and hydrophobic residues that specifies the recognition of discrete substrate motifs. This combinatorial recognition of polypeptides broadens TRiC specificity and may direct its unique ability to fold obligate substrates.

Antirecycling Agenda for Yeast

PAGE 1056

Yeast cells never re-establish polarity at the same site as was used in a previous cell division. Meitinger et al. shows that these sites, called cytokinesis remnants, are stably loaded with inhibitory polarity cues that prevent Cdc42-mediated growth and forstall premature replicative aging.

C-RACKing Open Viral Replication

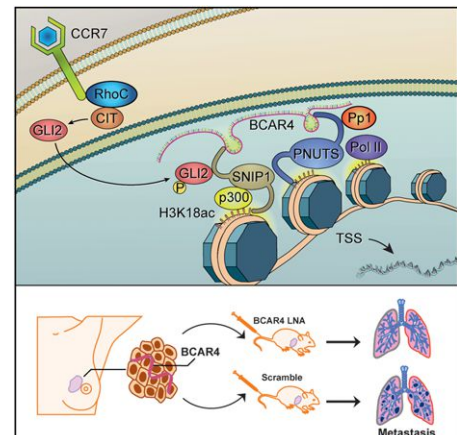
PAGE 1086

Majzoub et al. show that the ribosomal protein RACK1 plays a critical role in the translation of IRES-dependent viral RNAs but is not essential for cell viability and proliferation. These findings reveal that RACK1 has selective functions in mRNA translation and is target for the development of antivirals targeting IRES-dependent viruses, such as hepatitis C.

LncRNA Wires Up Cancer Metastasis

PAGE 1110

Xing et al. identify that the lncRNA BCAR4 contributes to breast cancer metastasis and elucidate how it triggers Hedgehog target genes in response to chemokine signals. The mechanism involves the coordinate binding of BCAR4 to two distinct transcriptional regulators and the direct and indirect effects of these interactions on oncogenic gene expression.



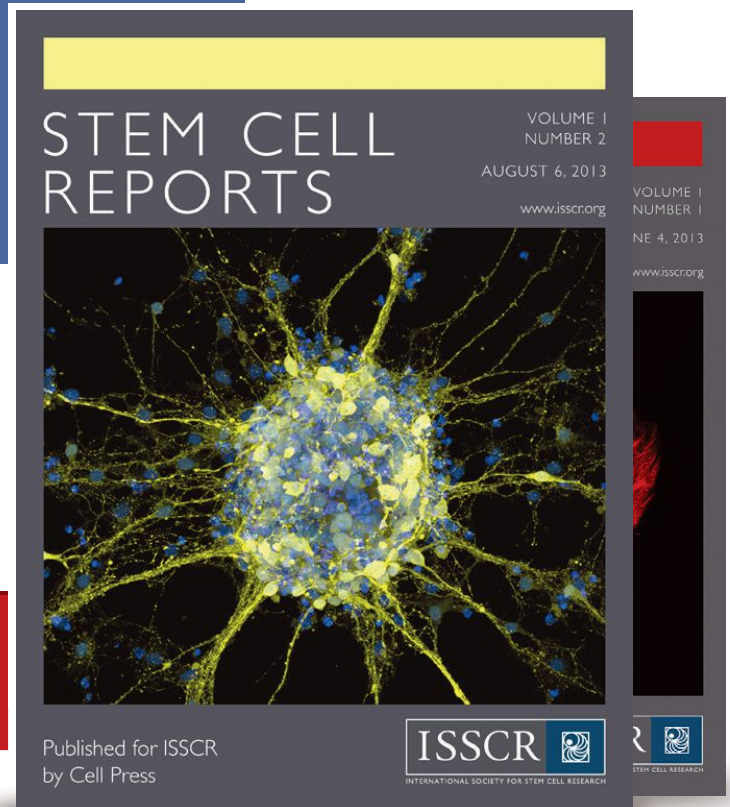
INTRODUCING STEM CELL REPORTS

EDITOR-IN-CHIEF

Christine Mummery, PhD,
Leiden University Medical Center

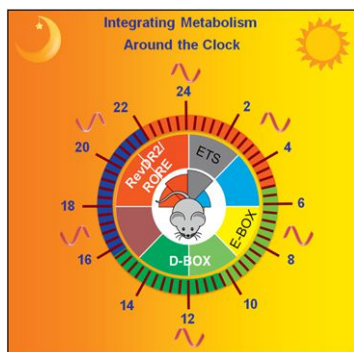
**CHECK OUT THE
LATEST ARTICLES ONLINE**

cell.com/stem-cell-reports



“ This open-access forum will
provide high visibility for your
stem cell research ”

Shinya Yamanaka, MD, PhD
ISSCR President and Nobel Laureate



Enhancers Work around the Clock

PAGE 1140

By examining enhancer RNAs that cluster in specific circadian phases, Fang et al. reveal the functional component of transcription factor cistromes from mouse liver. This analysis provides global insight into the partitioning of circadian gene expression, including the revelation that a single transcription factor controls opposing circadian phases.

Gag and RNA Get into an Open Relationship

PAGE 1096

The HIV-1 RNA-binding protein Gag orchestrates all steps of virion genesis, including membrane targeting and RNA recruitment into virions. Now, Kutluay et al. report that Gag-RNA-binding specificity dramatically and transiently changes during virion genesis, facilitating viral genome packaging.

Super-Enhancer Super-Inhibitor

PAGE 1126

MYC oncoproteins are thought to stimulate tumor cell growth and proliferation through amplification of gene transcription. Chipmuro et al. report that a covalent inhibitor of the kinase CDK7 induces regression in a mouse model of neuroblastoma through massive downregulation of MYC-induced transcriptional amplification by preferentially disrupting super-enhancer-associated genes.

A Matrix of Neo-Drug Interactions

PAGE 1168

The yeast *Cryptococcus neoformans* is the major fungal cause of meningitis as well as driver of mortality in HIV/AIDS patients. Brown et al. perform a large-scale investigation of chemical-genetic interactions in *C. neoformans* and exploit these data to delineate numerous genes involved in the biosynthesis of capsular polysaccharide, a central virulence trait, to identify the cell cycle as a key target of an antifungal compound, and to develop an algorithm that successfully predicts synergistic drug interactions.

Green Glutamine Sensor

PAGE 1188

Plants have long been thought to lack glutamine sensors. Chellamuthu et al. now show the plastid-localized PII signaling protein is a widely used glutamine sensor in the plant kingdom. Its short C-terminal extension forms a low-affinity glutamine-binding site, which is curiously lacking in Brassicacea, the family of flowering plants that includes the extensively studied model *Arabidopsis thaliana*.

Ribosomes Stop and Drop

PAGE 1200

During translation of messenger RNAs, ribosomes pause at some codons but not others. Subramaniam et al. combine experimental and computational approaches to show how nutrient status, reflected in tRNA aminoacylation, promotes or pauses translation elongation and show that aborted translation regulates ribosome density.

The Human Network

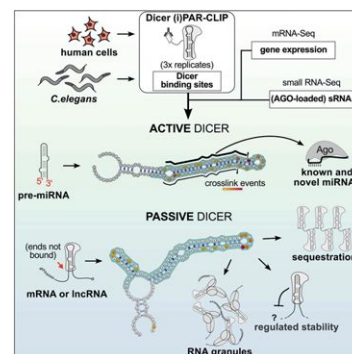
PAGE 1212

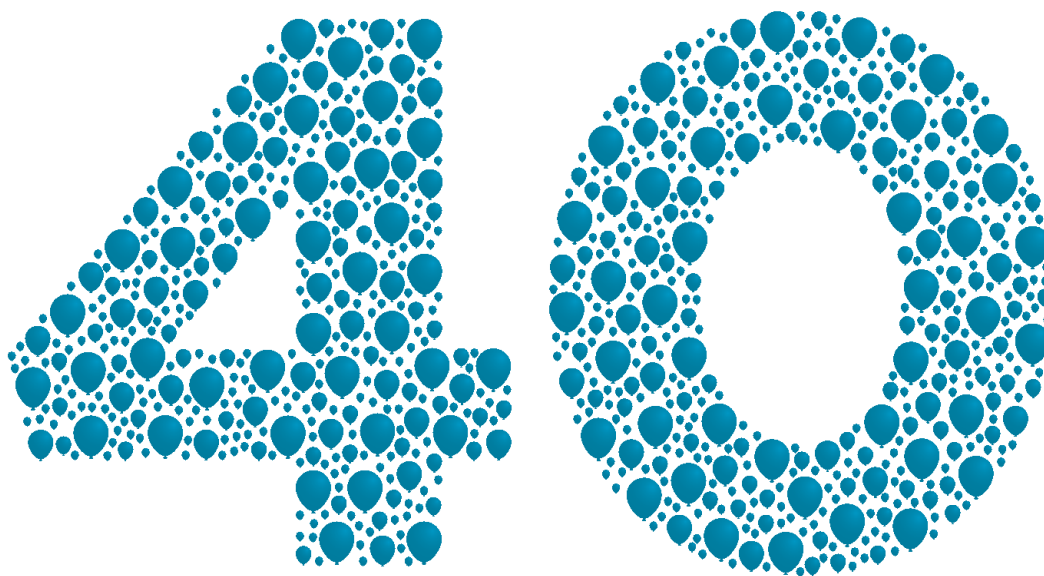
Rolland et al. create a systematic map of ~14,000 high-quality human binary protein-protein interactions to provide proteome-wide coverage of the human interactome network. The map uncovers significant interconnectivity between known and candidate cancer gene products, providing unbiased evidence for an expanded functional human disease landscape.

Dicer Is a Picky Eater

PAGE 1153

Dicer is known for its role in the biogenesis of eukaryotic small RNAs/microRNAs. Rybak-Wolf et al. now map transcriptome-wide Dicer-binding sites in human cells and *C. elegans* and identify hundreds of new miRNA, structural RNA, promoter RNA, and mitochondrial transcript targets. Interestingly, most Dicer-binding sites on mRNAs/lncRNAs are not significantly processed into small RNAs, but instead act to sequester Dicer.





YEARS TOGETHER



Join us throughout 2014 as we commemorate 40 years of publishing exciting biology. Cell-ebate with us – at Cell.com, in print issues, and at conferences – as we take stock of the journey we’ve shared since 1974.

www.cell.com/40/home

Cell

History Lessons from Bacteria

Microbes have been catapulted into the spotlight again by recent studies showing that uniquely modern phenomena, like jet lag and artificial sweeteners, can affect health by altering the microbiome (Suez et al., 2014; Thaïss et al., 2014). However, even as the nature of our current relationship with bacteria unfolds, DNA sequencing and proteomic approaches are giving us a fascinating glimpse into the history of human-bacterial interactions. By dusting off ancient human remains, recent work has uncovered some interesting findings about our ancestors' relationship with microbes.



Moche jar with fine line painting and figural seal on top. Moche culture, Peru, Phase III, 4th-6th century AD. Linden-Museum Stuttgart; photograph: Anatol Dreyer; Inventory-Number: 119020.

One example involves the spread of tuberculosis (TB) to the New World. *Mycobacterium tuberculosis* strains currently found in the Americas are related to those in Europe, suggesting European origins, but archeological evidence for tuberculosis in the New World prior to European contact suggests that TB was not spread through early settlers or traders. Another possible route is through human migration across the Bering land bridge during the Pleistocene era, but if this were the case, the dominance of the European lineage in the Americas would present yet another unexplained puzzle. How, then, did TB manage to get to the New World? A recent study led by Johannes Krause (Bos et al., 2014), whose previous work includes genome sequencing of ancient bacteria causing leprosy and the plague (Schuene-mann et al., 2013; Bos et al., 2011), sheds some light on this historical-epidemiological mystery.

The authors isolated *M. tuberculosis* DNA from 1,000 year old Peruvian human skeletons and compared the sequences to those previously identified from an 18th century Hungarian mummy (Chan et al., 2013), to modern human TB strains, and to related bacteria from a number of different animals. Surprisingly, they found that the Peruvian *M. tuberculosis* strains were not most closely related to the other human strains, as might have been expected. Instead, they clustered with bacterial strains found in seals populating the southern hemisphere. Yes, seals. Given that hunting likely afforded ample opportunity for contact between people and seals, this suggests a rather astonishing answer to how TB may have spread from Europe to the Americas: by hitching a ride across the ocean in marine mammals.

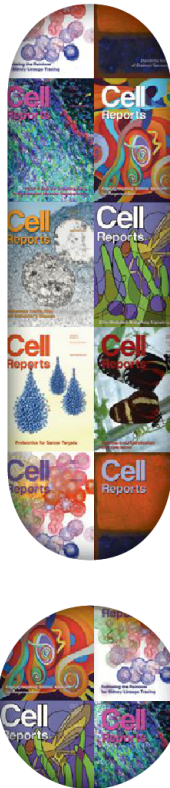
Not only has the sequencing of ancient DNA uncovered how a deadly pathogen may have colonized the New World long before Columbus did, but it's also giving us insights into ancient microbes that stuck much closer to home. The oral cavity is home to a rich community of bacteria, whose DNA can be encased and preserved within dental plaque for thousands of years. Since our ancestors didn't benefit from regular visits to the dental hygienist, this provides a source of material from which ancient oral microbiomes can be examined and has permitted work showing that with the advent of agriculture came a shift in the oral bacteria that cause gum disease (Adler et al., 2013). A study earlier this year, led by Enrico Cappellini, looks into the more recent past by sampling the dental plaque from German medieval skeletons (Warinner et al., 2014).

The study combines DNA sequencing and proteomic approaches to paint a picture of microbial pathogenicity, host immunity, and diet in the Middle Ages. Despite changes in diet and dental hygiene since that time, it seems that we share some of the same pathogenic oral bacteria with our medieval counterparts. In addition, the molecular machinery



Fossilized plaque on teeth of middle-aged man from Dalheim, Germany, ca. AD 1100. Photo credit: Christina Warinner.

Make your point with Cell Reports



Share your discoveries as you find them. Every breakthrough should get the attention it deserves. *Cell Reports* puts a spotlight on groundbreaking research with shorter, single-point stories.

Cell Reports is an open access journal that offers the quality, rigor, and visibility you've come to expect from Cell Press.

Do you have a new biological insight? Submit your paper to *Cell Reports*.

For more information www.cell.com/cell-reports

An open access journal with impact.
From Cell Press.

Cell
Reports

allowing bacteria to mount broad, low-level resistance to antibiotics was present even back then, as were host proteins involved in the innate immune response. Remarkably, some molecular remnants of food were also preserved in the plaque. Although leafy greens don't fossilize well, the study presents paleobiological evidence that the medieval diet included broccoli-related vegetables, among other food-stuffs. Given that diet can affect microbiome composition, the accessibility of ancient dietary habits to genomic and proteomic analysis may help us better understand how changing nutrition has impacted the microbiome throughout history.

As DNA sequencing technologies continue to advance, we'll likely get an even closer look at the complex ways in which human and microbial histories have intertwined. In addition, given that proteins tend to be more stable than DNA, proteomic approaches may open a window even farther into the past than we've been able to access until now. Apart from giving us a slightly apprehensive view of our cetacean friends, or a sense of common humanity with the ancients the next time we eat some broccoli, moving forward, these kinds of studies promise to give us a more complete picture of the effects of microbes on human health, both past and present.

REFERENCES AND RECENT RELATED PAPERS

- Adler, C.J., Dobney, K., Weyrich, L.S., Kaidonis, J., Walker, A.W., Haak, W., Bradshaw, C.J., Townsend, G., Sołtysiak, A., Alt, K.W., et al. (2013). *Nat. Genet.* 45, 450–455, e1.
- Bos, K.I., Schuenemann, V.J., Golding, G.B., Burbano, H.A., Waglechner, N., Coombes, B.K., McPhee, J.B., DeWitte, S.N., Meyer, M., Schmedes, S., et al. (2011). *Nature* 478, 506–510.
- Bos, K.I., Harkins, K.M., Herbig, A., Coscolla, M., Weber, N., Comas, I., Forrest, S.A., Bryant, J.M., Harris, S.R., Schuenemann, V.J., et al. (2014). *Nature* 514, 494–497.
- Chan, J.Z., Sergeant, M.J., Lee, O.Y., Minnikin, D.E., Besra, G.S., Pap, I., Spigelman, M., Donoghue, H.D., and Pallen, M.J. (2013). *N. Engl. J. Med.* 369, 289–290.
- Schuenemann, V.J., Singh, P., Mendum, T.A., Krause-Kyora, B., Jäger, G., Bos, K.I., Herbig, A., Economou, C., Benjak, A., Busso, P., et al. (2013). *Science* 341, 179–183.
- Suez, J., Korem, T., Zeevi, D., Zilberman-Schapira, G., Thaïs, C.A., Maza, O., Israeli, D., Zmora, N., Gilad, S., Weinberger, A., et al. (2014). *Nature* 514, 181–186.
- Thaïs, C.A., Zeevi, D., Levy, M., Zilberman-Schapira, G., Suez, J., Tengeler, A.C., Abramson, L., Katz, M.N., Korem, T., Zmora, N., et al. (2014). *Cell* 159, 514–529.
- Warinner, C., Rodrigues, J.F., Vyas, R., Trachsel, C., Shved, N., Grossmann, J., Radini, A., Hancock, Y., Tito, R.Y., Fiddyment, S., et al. (2014). *Nat. Genet.* 46, 336–344.

Cindy Lu



slas
2015

4th annual
CONFERENCE & EXHIBITION

COME TRANSFORM RESEARCH AT SLAS2015

february / **seven - eleven**

WASHINGTON, DC / **SLAS2015.ORG**
WALTER E. WASHINGTON CONVENTION CENTER

Join more than 5,000 of the brightest, most innovative minds vested in scientific automation at SLAS2015 as they come together to learn, share and explore how to transform scientific research using technology. Scientists, engineers, researchers, business leaders and technology providers from industry, academia and government backgrounds will benefit from 132 peer-selected podium presentations comprising 33 unique scientific sessions, hundreds of posters, plentiful networking opportunities and access to the latest technologies from 300+ exhibitors from around the world. Participating at SLAS2015 is an invaluable investment that pays significant dividends toward your personal and organizational goals, as well as your professional success.

REGISTER NOW AT SLAS2015.ORG

SLAS2015 Scientific Program Crosses a Diversity of Industries and Interests

- Drug Discovery and Development
- Informatics
- Clinical Diagnostics
- Technology Transfer
- Laboratory Instrumentation
- Synthetic Chemistry
- Healthcare

SLAS2015 Educational Tracks

- Assay Development & Screening
- Automation & High-Throughput Technologies
- Bioanalytical Techniques
- Biomarker Discovery & Applications
- Drug Target Strategies
- Informatics
- Micro/Nano Technologies

KEYNOTE PRESENTERS



Francis Collins
Director, U.S.
National Institutes
of Health (NIH)



Donald E. Ingber
Harvard University



Laurie Garrett
Renowned Author
and Authority
on Global
Health Issues



Scan the code with your mobile device
to go directly to **SLAS2015.org**



slas

SOCIETY FOR
LABORATORY AUTOMATION
AND SCREENING

Credibility and Reproducibility

Credibility is everything for science, and it is built over time in both obvious and subtle ways. It is how we interact with colleagues and collaborators. It is how generously and openly we share reagents and how we mentor students and postdocs. It is how we review each other's papers, and it is how we credit others' work. It is the way we educate and inform the public that funds us. It is the way we document and store our data. And it is the rigor, transparency, and attention we invest in designing, conducting, and reporting experiments. Without credibility, others can't/won't build on our work, and as a result, the pace of scientific advance is slowed. Most importantly, science contaminated with a lack of credibility is a house with crumbling walls that engenders little trust and provides minimal value to our global society, present and future.

But everyone reading this already knows the importance of credibility in science, so why are we discussing it here? Within the last 12 months the reproducibility of science, a lynchpin of credibility, has come under intense scrutiny, both from the NIH (Nature 505, 612–613) and other government funding bodies, as well as in the lay (The Economist, October 19, 2013, 23–28) and scientific press (Nature 483, 531–533—though many of these reports themselves would benefit from greater transparency in reporting and still require robust demonstrations of reproducibility). Hearing the word “reproducibility,” most of us think immediately of fraud or data and image manipulation, but it is much broader than that. Many of the current concerns about reproducibility, particularly the successful scalability of preclinical data into robust drug targets for treating human disease, are focused on the rigor of the experimental design (inclusion of all appropriate controls, blinded experimental conditions, gender balance in experimental populations, a priori determination of n's and statistical power, appropriate statistical analyses, etc.) and on complete transparency in reporting of these parameters and all collected data (for a recent Perspective on this topic, see Neuron 84, 572–581.)

In June of this year Francis Collins, NIH Director, Marcia McNutt, Editor-in-Chief of *Science*, and Philip Campbell, Editor-in-Chief of *Nature*, organized a meeting of journal editors and other contributors to collaborate on approaches to ensuring and improving reproducibility. Maximizing reproducibility clearly is an initiative involving many stakeholders, with scientists front and center and funding bodies, universities, journals, pharmaceutical and biotech companies, patient advocacy groups, and society at large all taking a leading role as well. Out of the discussions at this meeting came a set of recommendations for how journals and journal editors can do their part. The main focus of the guidelines is to ensure rigorous experimental design and transparency in reporting the specifics about how experiments were performed and how data were collected and analyzed. Cell Press participated in the meeting and is a signatory on the recommendations that were recently posted (<http://www.nih.gov/about/reporting-preclinical-research.htm>). Many of the items in the guidelines *Cell* and its sister journals are already

doing and have been doing for quite some time (providing space for lengthy methods sections in print and unlimited supplemental methods online, requiring the sharing of reagents as a condition of publication, providing a forum for refutation in our Matters Arising format, requiring authors to clearly state their statistical measures.) Other items in the guidelines, like developing a way to facilitate clear reporting in the paper of details about how experiments were designed and performed, will be valuable additions to what we already do. Journals are encouraged to adopt a checklist of specific reporting criteria as a standard form for authors to complete and editors and/or reviewers to verify. While we at *Cell* and the other Cell Press journals are not yet sure that an author checklist per se will be the most effective implementation for our authors, reviewers, and readers, we do wholly embrace the importance of the goals of the guidelines and will be taking steps to adapt our editorial processes and author instructions to ensure consistent standards for appropriate experimental design and transparency in reporting. For example, *Developmental Cell* has recently introduced supplemental protocols, where authors of a paper with noteworthy, new, or particularly challenging methods are encouraged to provide a detailed protocol in a separate supplemental PDF. We view these steps as an important part of the value that we add through the editorial and peer review process.

Enhanced attention to these elements will also help protect the authors' credibility. With increased clarity about how experiments were performed and collected, editors, authors, and reviewers will all be better able to spot and rectify concerns before the paper is published, hopefully reducing the number of corrections and retractions required postpublication. To this end, *Cell* and our sister journals are also introducing an image screening process to help ensure adherence to community standards as outlined in our data processing policies. More and more, we are finding that the concerns that arise regarding published data are often the result of avoidable errors. For example, copying and pasting the same image into two different figures or failing to indicate where lanes of a gel have been spliced together. (Oddly, the most pervasive challenges to published data we see at *Cell* relate to loading controls. There seems to be some misalignment among scientists regarding the importance and meaning of the actin bands in a standard western blot.) When potential problems are brought to our attention by a concerned reader, we ask the authors to provide us with the original unprocessed data, together with a detailed explanation of how they conducted the experiment. Most of the time, we can see from the raw data that the problems have been introduced through simple mistakes and can be addressed with an erratum. But a scientific literature peppered with corrections does not build credibility, and worse still is when the avoidable errors are sufficiently extensive that they undermine the reliability of the entire body of work and necessitate a retraction. So, as we at *Cell* invest in checking figures and working with authors to fix any correctable mistakes before the paper is published, we ask

that authors renew their focus on preparing their manuscripts and reviewing the final figures with the same attentive eyes their readers will. By combining enhanced clarity of reporting as recommended by the new guidelines with prepublication image screening, our intent is to ensure that every paper we publish meets not only the highest standards of interest and importance but also of credibility and reproducibility.

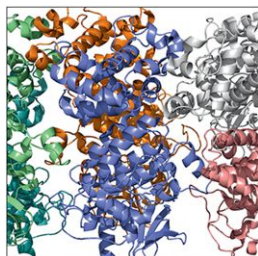
With increased vigilance from authors, funders, and journals and attention to standards for experimental design and accurate careful reporting, we will collectively increase the public trust and support for research and build a stronger pipeline for converting our understanding of the basic processes and mechanisms of biology into improved diagnostics, treatments, and potentially cures for the myriad of global health challenges.

Emilie Marcus

<http://dx.doi.org/10.1016/j.cell.2014.11.016>

What's Your Favorite Crystal Structure?

Signaling Helix

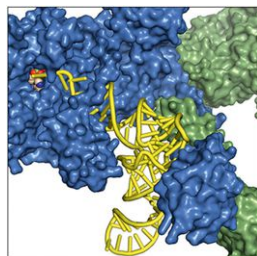


Hao Wu
Harvard Medical School

I stumbled into the crystal structure of Ire1 a few years ago when my lab had just solved the structures of several helical signaling oligomers formed by death domains. I was curious if others had observed similar structures in their systems and whether these higher-order oligomers might tell us something new about molecular mechanisms in signal transduction. Ire1 is a receptor critical for the unfolded protein response (UPR) in the endoplasmic reticulum (ER). It contains an ER luminal region that senses aberrantly folded proteins and a cytoplasmic region with a kinase domain and an RNase domain. Upon activation, the Ire1 RNase mediates nonconventional mRNA splicing to enable expression of transcription factors that activate the UPR.

The crystal structure of the Ire1 cytosolic domain reveals a striking oligomeric assembly, corroborating the observed activation of the Ire1 RNase in solution by oligomerization (Nature [2009], 457, 687–693). Symmetric back-to-back Ire1 dimers arrange into a filament through helical symmetry. Each kinase domain extends its activation loop to the next kinase domain, resembling a *trans*-autophosphorylation reaction. Remarkably, the RNase active site that is disordered in a dimer structure becomes ordered in the oligomer to create the mRNA-binding pocket. The oligomerization-promoted allosteric activation of the RNase made perfect sense to me, and an analogous mechanism may bring about kinase and caspase activation in the context of innate immune signaling. A picture is worth a thousand words.

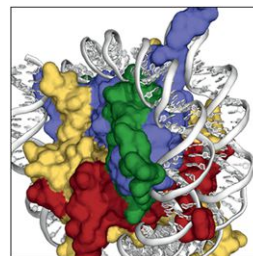
A Single Base Pair



Shigeyuki Yokoyama
RIKEN and The University of Tokyo

Reading the genetic code, which describes the rules relating amino acids to codons, relies on strict molecular recognition of both amino acids and tRNAs by aminoacyl-tRNA synthetases. Aminoacyl-tRNA synthetases must recognize key nucleotides far from the 3' terminal adenosine that is ligated to the amino acid. The structures of alanyl-tRNA synthetase (AlaRS) in complex with tRNA^{Ala} revealed how AlaRS selects tRNA^{Ala} depending just on a single base pair G3·U70 (Nature [2014], 510, 507–511) and thereby unraveled what had been a mystery since 1988. A variant of tRNA^{Ala} with A3·U70 can bind to AlaRS with nearly the same affinity, but aminoacylation is about hundred times slower than wild-type with G3·U70. The AlaRS·tRNA structure of the A3·U70 variant showed that the CCA sequence is bound in a site far removed from the catalytic site, which traps the 3'-adenosine. Moreover, G3·U70 is able to control the direction of the 3'-CCA. It is surprising that such high translation specificity is achieved by this unprecedented mechanism. The finding of this nonproductive binding mechanism may cause a paradigm shift in enzymology of not only aminoacyl-tRNA synthetases, but also other highly specific enzymes. Furthermore, rational engineering of aminoacyl-tRNA synthetases toward novel specificities for tRNAs may become possible in the field of synthetic biology, utilizing unnatural amino acids and base pairs in the central dogma.

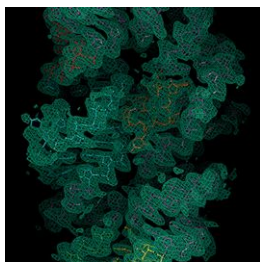
That Exact Moment



Karolin Luger
Colorado State University and HHMI

Everybody's all-time favorite structure has to be the DNA double helix because it exemplifies everything a high-impact structure should do: be aesthetically pleasing and explain complex biology. After this obvious choice, with full awareness of my bias, my favorite structure is the nucleosome. Obviously, the fact that I participated for 8 long years in the arduous process leading to a high-resolution model of the repeating unit of eukaryotic chromatin affects my choice. But on an objective scale, the nucleosome's beauty lies in its inherent 2-fold symmetry, its aesthetically pleasing proportions, and the way in which intricately intertwined histone helices generate a ramp that gently guides the DNA double helix into a superhelix. More importantly, the structure beautifully explains its function in organizing the entire eukaryotic genome: the histone octamer holds the DNA with just enough force to bend it around its perimeter yet allows its dissociation for key biological processes to take place. Its modular design lends itself to a dynamic response to regulate transcription, replication, and repair. Finally, the histone tails, the sites of numerous posttranslational modifications, are freely accessible for signaling and for promoting internucleosomal interactions (Nature [1997], 389, 251–260). The entire structure had to be manually fit into the electron density map, and so there was no single moment at which I first "saw" the nucleosome. But when I projected the structure for the first time at a Cold Spring Harbor meeting, it was greeted with a collective gasp. I remember that exact moment to this day.

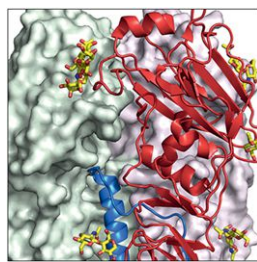
The 3D Beauty of RNA



Jennifer Doudna
UC Berkeley and HHMI

As a graduate student in the late 1980s, with a backdrop of Madonna tunes and post-punk fashion, I was fascinated by the hypothesis that RNA played a starring role during early evolution. Perhaps this agile molecule had both stored genetic information *and* catalyzed the chemical reactions that ensured its replication, enabling evolution to get going. Thinking that the three-dimensional structure of RNA was key to understanding the function and possible origins of ribozymes, I worked with then-student Jamie Cate to determine the crystal structure of the P4-P6 domain of the *Tetrahymena* self-splicing ribozyme (Science [1996]. 273, 1678–1685). I'll never forget my sense of wonder as the first electron density maps revealed the beauty of the RNA structure: its coiled helices were dramatically bent to form a distinctive swan's-neck shape, with magnesium ions nestled in the center to hold together phosphate backbones that would otherwise spring apart due to electrostatic repulsion. At 2.8 Å resolution, the 160 nucleotide P4-P6 structure was brimming with exciting details. Recurrent motifs, including the tetraloop-receptor interaction, the A-A platform, A-minor contacts, and ribose zippers, were subsequently observed in complex RNAs ranging from the ribosome to riboswitches, setting the stage for many functional studies and mechanistic insights. But for me, the first electrifying moments of seeing the P4-P6 RNA structure emerge from initially noisy electron density remains a standout in my career.

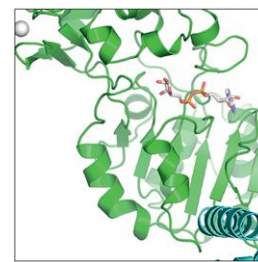
Flu Escapes Immunity



Ian Wilson
The Scripps Research Institute

The crystal structure of influenza virus hemagglutinin (HA) in 1981 revolutionized our thinking on how animal viruses bind and enter host cells and how they evade our immune system (Nature [1981]. 289, 366–373 and 374–378). At the time, the trimeric HA glycoprotein from the 1968 Hong Kong pandemic virus was considered a very large protein (200 kDa) for structural studies. Given its composition of 25% glycosylation by weight, novel methods were required to solve its structure, using only one heavy atom derivative with 3-fold symmetry averaging and solvent flattening (>80% solvent). Its now classic fold was composed of a jelly roll globular head for receptor binding and a largely helical stem housing the fusion machinery. Mapping of natural mutations from 1968 to 1977 on the HA structure uncovered four major antigenic sites, elucidating how the virus can escape immune recognition and why we need a seasonal vaccine. It was a stunning moment as the structure came to life and explained a huge amount of influenza virology and biology. The recent discovery of broadly neutralizing antibodies to flu has now uncovered key sites of vulnerability on the HA of this virus and has given hope for structure-based design of a universal vaccine. Remarkably, 33 years after the first structure of HA, fascinating insights on this viral glycoprotein are still being revealed.

Visualize Longevity Modulator



Rui-Ming Xu
Chinese Academy of Sciences

The discovery in 2000 that Sir2 and related proteins (sirtuins) are NAD-dependent histone/protein deacetylases was a landmark achievement following more than a decade's effort of genetic and biochemical characterization of yeast Sir2 function in transcriptional silencing. Sirtuins are found in all three kingdoms, but their sequences give no hints of their 3D structures and biochemical mechanisms. The first structures of Sirtuins, an archaeal protein in complex with NAD and an apo form of human SIRT2, revealed a prominent NAD-binding Rossmann-fold domain (Cell [2001]. 105, 269–279). NAD was bound in an inverted orientation compared to the canonical binding mode. This, together with the presence of a zinc-binding domain, immediately pointed toward the locations of the acetyllysine-binding channel and the active sites for deacetylation and nicotinamide cleavage. Now the catalytic mechanisms have been worked in considerable detail with the help of many structural studies. In addition, the mechanisms regulating sirtuin activities have begun to emerge, with the determination of the Sir4-bound structure of Sir2, which appears to lock the regulatory and catalytic domains of Sir2 in a defined configuration to allow productive deacetylation reactions. Mammalian SIRT1 shares a similar domain structure with Sir2, and SIRT1 has been implicated in many physiological and pathological processes, including aging, cancer, and cardiovascular and metabolic diseases. There has been keen interest in developing small-molecule modulators for SIRT1, an area where X-ray crystallography can make further contributions.

Accelerating HIV-1 Vaccine Efficacy Trials

Dan H. Barouch^{1,2,*} and Nelson L. Michael^{3,*}

¹Center for Virology and Vaccine Research, Beth Israel Deaconess Medical Center, Harvard Medical School, Boston, MA 02215, USA

²Ragon Institute of MGH, MIT, and Harvard, Cambridge, MA 02139, USA

³U.S. Military HIV Research Program, Walter Reed Army Institute of Research, Silver Spring, MD 20910, USA

*Correspondence: dbarouch@bidmc.harvard.edu (D.H.B.), nmichael@hivresearch.org (N.L.M.)

<http://dx.doi.org/10.1016/j.cell.2014.10.046>

Despite major advances in HIV-1 therapeutics and prevention strategies, the development of a safe and effective prophylactic HIV-1 vaccine will likely be critical for ending the global HIV-1 epidemic. Yet only four HIV-1 vaccine concepts have been tested for clinical efficacy over the past 30 years. In this Commentary, we describe key hurdles facing the HIV-1 vaccine development field and outline strategies to accelerate efficacy evaluation of novel HIV-1 vaccine candidates.

Current State of the HIV-1 Vaccine Field

Despite the urgent need for a globally effective prophylactic HIV-1 vaccine, only four HIV-1 vaccine concepts have been tested in six clinical efficacy trials to date (Table 1). These concepts have included (1) Env gp120 proteins, (2) recombinant adenovirus serotype 5 (rAd5) vectors, (3) canarypox (ALVAC) vectors with gp120 boosts, and (4) DNA vaccines with a rAd5 boost. The first efficacy studies evaluated monomeric HIV-1 envelope (Env) gp120 protein vaccines with alum adjuvant and were tested in two phase III vaccine trials. These vaccines failed to prevent HIV-1 acquisition in men who have sex with men (MSM) and high-risk women in the United States and Europe (Vax004) (Flynn et al., 2005), as well as in injection drug users in Thailand (Vax003) (Pitisuttithum et al., 2006).

A rAd5 vector-based vaccine expressing the HIV-1 internal proteins *gag/pol/nef* was then tested in the Step (HVTN 502) and Phambili (HVTN 503) phase IIb trials. The Step trial, which was conducted in MSM and high-risk women in the Americas, Caribbean, and Australia, was stopped for futility to block HIV-1 acquisition (Buchbinder et al., 2008). Subsequent analyses suggested an increase in HIV-1 acquisition in vaccinees, particularly in the subgroup of uncircumcised men who were seropositive at baseline for Ad5. This finding cast a pall over the HIV-1 vaccine development field and led to increased research emphasis on the potential importance of vector-specific immune responses. The Phambili study

tested the same vaccine in high-risk heterosexuals in South Africa and was stopped during its enrollment phase shortly after the Step results were announced (Gray et al., 2011). Unblinded follow-up of Phambili participants suggested a very late effect of excess HIV-1 infections in heterosexual male vaccinees without a clear mechanism of action (Gray et al., 2014). Concordant with the lack of efficacy observed in these clinical trials, preclinical studies similarly demonstrated that analogous rAd5 vectors expressing *gag/pol/nef* from the related simian immunodeficiency virus (SIV) afforded no protection against acquisition of infection following mucosal SIV challenges in rhesus monkeys (Reynolds et al., 2012).

The third vaccine concept involved priming with a canarypox vector (ALVAC) expressing the HIV-1 antigens *gag/pol/env* and boosting with the same gp120 protein subunits that were used in the Vax003 study. The RV144 study was conducted in a low-incidence, mostly heterosexual population in Thailand and demonstrated vaccine efficacy of 31% at 42 months (Rerks-Ngarm et al., 2009). Efficacy was 60% at 12 months, indicative of an early protective effect that waned over time. Subsequent analyses demonstrated that the risk of HIV-1 infection correlated inversely with antibodies directed against the first and second HIV-1 Env variable regions (V1V2) and correlated directly with Env-specific IgA antibodies (Haynes et al., 2012). Additional analyses suggested that HIV-1 infection risk also inversely correlated most clearly with V2-specific antibodies of the IgG3 isotype and non-neutralizing

functional activity. Furthermore, a molecular sieve analysis showed immune selection pressure on specific V2 amino acids in vaccinees (Rolland et al., 2012). Consistent with the clinical results, modest protective efficacy was also observed with analogous ALVAC/gp120 vaccines against mucosal SIV challenges in rhesus monkeys.

The fourth vaccine concept that was tested involved priming with DNA vaccines expressing *gag/pol/nef/env* and boosting with rAd5 vectors expressing *gag/pol/env* in the HVTN 505 study, which was a phase IIb study conducted in MSM in the Americas. Importantly, preclinical data showed that this vaccine afforded partial protection against low-stringency SIV challenges (strain SIVsmE660) in rhesus monkeys but failed to protect against high-stringency SIV challenges (strain SIVmac251) (Letvin et al., 2011). HVTN 505 was halted at its first interim efficacy analysis for futility to protect against HIV-1 acquisition or lower HIV-1 viral RNA in breakthrough infections (Hammer et al., 2013). These data strongly suggest that preclinical studies of HIV-1 vaccines should be evaluated exclusively in stringent preclinical challenge models.

Future HIV-1 Vaccine Efficacy Studies

Several HIV-1 vaccine candidates are expected to be evaluated in clinical efficacy studies in the next few years. The Poxvirus-Protein Public Private Partnership ("P5") is a collaborative group that has been formed to build on the results of the RV144 trial and to test the identified

Table 1. Clinical HIV-1 Vaccine Efficacy Trials

Study	Vaccines	Phase	Risk Group	HIV Incidence per 100 Person-Years	Location	Result
Vax003	AIDSVAX B/E gp120 in alum	III	injecting drug users	3.40%	Thailand	no vaccine efficacy
Vax004	AIDSVAX B/B gp120 in alum	III	high-risk women and MSM	2.60%	United States, Europe	no vaccine efficacy
HVTN 502 Step	MRKAd5 HIV-1 gag/pol/nef B	IIb	high-risk women and MSM	3.00%	United States	halted at interim analysis for futility; early transient increased infection in vaccinees
HVTN 503 Phambili	MRKAd5 HIV-1 gag/pol/nef B	IIb	high-risk heterosexual men and women	3.70%	South Africa	no vaccine efficacy; late increased HIV infection in unblinded male vaccinees
RV144	ALVAC-HIV vCP1521, AIDSVAX B/E rgp120 in alum	III	community risk heterosexual men and women	0.28%	Thailand	31.2% efficacy at 42 months as primary endpoint; 60% efficacy at 12 months
HVTN 505	DNA, rAd5 (A, B, C)	IIb	circumcised MSM without pre-existing Ad5 antibodies	1.80%	United States	halted at interim analysis for futility

MSM, men who have sex with men; Ad5, adenovirus serotype 5.

V2 correlate of risk in a new series of HIV-1 vaccine efficacy trials in sub-Saharan Africa and Thailand. For studies in Africa, new ALVAC and gp120 vaccine products with HIV-1 clade C antigens are currently being manufactured. ALVAC vectors expressing clade C antigens and MF59-adjuvanted gp120 subunits are planned for a phase III licensure study in sub-Saharan Africa. Other strategies, including NYVAC and DNA-NYVAC priming prior to gp120 boosting, are planned for evaluation in non-licensure track phase IIb efficacy studies. High-risk MSM cohorts are also being explored in Thailand for further efficacy testing.

Additional HIV-1 vaccine candidates are also being developed for efficacy testing. A recombinant adenovirus serotype 26 (rAd26) prime expressing *gag/pol/env* with a modified vaccinia Ankara (MVA) boost expressing the same antigens has shown substantial protection against both SIVmac251 and SHIV-SF162P3 challenges in rhesus monkeys (Barouch et al., 2012, 2013). Boosting with an Env gp140 trimer appears to improve this observed protective efficacy in monkeys (D.H.B., unpublished data). A prototype rAd26 vector expressing *env* was also recently shown to be safe and immunogenic in humans with no evidence for activation of total or vector-specific CD4+ T cells in colorectal mucosa (Baden

et al., 2014). A multivalent Ad26/MVA vaccine expressing HIV-1 “mosaic” *gag/pol/env* immunogens designed for optimal coverage of global virus diversity, together with a stable Env gp140 trimer, are expected to enter clinical trials later this year. These Ad26/MVA and Ad26/gp140 vaccines are currently under consideration for advancement into efficacy testing.

Other promising vaccine candidates that have entered early phase clinical trials include priming with DNA vaccines and boosting with MVA vectors. In addition, early phase clinical trials are planned with cytomegalovirus (CMV) vectors, which showed substantial virologic control and possible clearance of SIVmac251 in approximately half of vaccinated monkeys following challenge (Hansen et al., 2013).

Why So Few HIV-1 Vaccine Efficacy Studies?

Why have so few HIV-1 vaccine efficacy trials been conducted to date for a problem of such global importance? Factors include the enormous scientific challenges in designing a vaccine for a highly variable virus that integrates in the host genome, rapidly establishes latency, and effectively evades both humoral and cellular immune responses. Although certain correlates of risk were identified

in the RV144 study, it is currently not clear whether these are true mechanistic correlates of protection or whether they will prove generalizable beyond the RV144 study. In addition, although nonhuman primate challenge models have proven informative, they do not yet represent validated animal models that are necessarily predictive of clinical efficacy trials. As a result, large, complex, iterative clinical efficacy studies are required to show the efficacy of candidate HIV-1 vaccines in humans. Moreover, the primary endpoint of such studies has to be incident new HIV-1 infections in large populations of high risk individuals. Such studies are expensive and logistically challenging, and thus they pose substantial risk to the pharmaceutical industry that traditionally drives advanced clinical development of vaccines.

Accelerating Clinical Efficacy Trials

Accelerating efficacy testing of novel and promising HIV-1 vaccine candidates will be vital for the field. In the absence of a validated and generalizable immune correlate of protection, only carefully designed clinical efficacy trials can determine whether a HIV-1 vaccine works in humans. The results of such trials, regardless of their outcomes, will have major impacts on the HIV-1 vaccine field, and they will lead to immediate prioritization

and deprioritization of vaccine candidates and strategies. Studies that show partial protective efficacy will also refine our understanding of immune correlates of protection. The efficacy trials conducted to date have had surprising outcomes that have been discordant with the expectations of experts in the field, and thus the current state of knowledge is inadequate to predict the results of any such efficacy trials with certainty. To accelerate efficacy testing of next-generation HIV-1 vaccine candidates, increased industry involvement, mobilization of resources, expansion of the current vaccine pipeline, and robust preclinical challenge studies will likely prove critical.

Increased industry involvement would be highly desirable for the HIV-1 vaccine field, particularly for advanced clinical development. Industry provides unique expertise in terms of manufacturing, regulatory affairs, and product development, as well as downstream licensure capacity to produce and to deliver a vaccine in the event that efficacy trials are successful. Currently, industry involvement has been relatively modest for the reasons described previously. Both the NIH and the Bill & Melinda Gates Foundation have programs that actively support academic-industry partnerships, and these programs should be continued and expanded. Engaging and encouraging industry involvement at the earliest stages in vaccine development may interest a potential industry partner in a particular vaccine platform and may lead to increased involvement for advanced clinical development should the scientific rationale prove compelling. Early involvement of industry may also be critical, as companies may need the use of particular raw materials, cell lines, vectors, or manufacturing technologies to be consistent with their internal platforms or processes.

The size, cost, and logistic complexity of HIV-1 vaccine efficacy trials are substantial. Thus, in addition to the major current investments from the NIH, Gates Foundation, and multiple other organizations, mobilization of new resources would greatly accelerate HIV-1 vaccine efficacy studies. Three potential sources for new funding include industry, governments, and additional philanthropy. Pharmaceutical companies may provide direct

support for development activities, including clinical efficacy trials, if they are sufficiently interested in a particular vaccine product. Additional investment of governments worldwide would be highly enabling for the field. Finally, new philanthropic funding will also accelerate the development of an HIV-1 vaccine, as exemplified by the Ragon Institute of MGH, MIT, and Harvard.

Accelerating clinical efficacy trials of HIV-1 vaccine candidates also requires multiple distinct and scientifically promising vaccine candidates in preclinical and early phase clinical studies to be ready for efficacy testing. It is therefore essential to maintain and expand a diverse portfolio of vaccine concepts. For example, novel Env immunogens are being developed by multiple groups to elicit broadly neutralizing antibodies as well as to optimize functional non-neutralizing antibodies; a series of new vectors have been explored that expand the breadth and efficacy of virus-specific T cell responses; and potential global antigens have been developed that begin to address the challenge of global virus diversity. A robust pipeline of new concepts and fresh perspectives will also require the engagement and encouragement of young and early career investigators, particularly those from the developing world.

Expanding preclinical efficacy studies will also help to support the rationale for clinical efficacy trials. Although the ability of nonhuman primate challenge studies to predict the outcomes of clinical efficacy trials still remains uncertain, stringent SIV and SHIV challenges in rhesus monkeys represent the most robust model for assessing vaccine candidates prior to clinical efficacy trials. Such preclinical challenge studies should therefore be expanded, particularly for vaccine candidates under consideration for efficacy trials. When clinical efficacy data with these vaccines become available, the clinical results can then be used to refine and to improve the preclinical models.

Perspectives

HIV-1 vaccine development will likely be an iterative process. Robust basic research must continue but needs to be matched with clinical efficacy testing of

promising new vaccine candidates. Information learned from each rigorous efficacy trial will be pivotal and will provide clear directions for the field. A more detailed understanding of immune correlates of protection will also be obtained from these clinical efficacy trials and may, ultimately, reduce the need to conduct large studies for each new vaccine concept. However, at the present time, there is no way to determine whether a vaccine will prevent HIV-1 infection in humans other than clinical efficacy studies in which incident new HIV-1 infections are the primary endpoint.

There are numerous reasons for optimism in the HIV-1 vaccine field. The RV144 study showed that an HIV-1 vaccine is possible, and several novel vaccine candidates have demonstrated unprecedented efficacy in stringent nonhuman primate challenge studies. Our basic understanding of HIV-1-specific humoral and cellular immunity has expanded considerably, and preclinical and clinical immune correlates of protection have been identified in certain contexts. Increasing the momentum to accelerate the conduct of efficacy trials will substantially accelerate the development of a safe and effective HIV-1 vaccine, which will presumably be required to control the global HIV-1 pandemic.

ACKNOWLEDGMENTS

The authors thank Bill Snow for insights and comments. The views expressed are those of the authors and should not be construed to represent the positions of the U.S. Army or the Department of Defense.

REFERENCES

- Baden, L.R., Liu, J., Li, H., Johnson, J.A., Walsh, S.R., Kleinjan, J.A., Engelson, B.A., Peter, L., Abbink, P., Milner, D.A., Jr., et al. (2014). J. Infect. Dis. Published online August 26, 2014. <http://dx.doi.org/10.1093/infdis/jiu485>.
- Barouch, D.H., Liu, J., Li, H., Maxfield, L.F., Abbink, P., Lynch, D.M., Iampietro, M.J., SanMiguel, A., Seaman, M.S., Ferrari, G., et al. (2012). *Nature* 482, 89–93.
- Barouch, D.H., Stephenson, K.E., Borducchi, E.N., Smith, K., Stanley, K., McNally, A.G., Liu, J., Abbink, P., Maxfield, L.F., Seaman, M.S., et al. (2013). *Cell* 155, 531–539.
- Buchbinder, S.P., Mehrotra, D.V., Duerr, A., Fitzgerald, D.W., Mogg, R., Li, D., Gilbert, P.B., Lama, J.R., Marmor, M., Del Rio, C., et al.; Step

- Study Protocol Team (2008). *Lancet* 372, 1881–1893.
- Flynn, N.M., Forthal, D.N., Harro, C.D., Judson, F.N., Mayer, K.H., and Para, M.F.; rgp120 HIV Vaccine Study Group (2005). *J. Infect. Dis.* 191, 654–665.
- Gray, G.E., Allen, M., Moodie, Z., Churchyard, G., Bekker, L.G., Nchabeleng, M., Mlisana, K., Metch, B., de Bruyn, G., Latka, M.H., et al.; HVTN 503/Phambili study team (2011). *Lancet Infect. Dis.* 11, 507–515.
- Gray, G.E., Moodie, Z., Metch, B., Gilbert, P.B., Bekker, L.G., Churchyard, G., Nchabeleng, M., Mlisana, K., Laher, F., Roux, S., et al.; HVTN 503/Phambili study team (2014). *Lancet Infect. Dis.* 14, 388–396.
- Hammer, S.M., Sobieszczyk, M.E., Janes, H., Karuna, S.T., Mulligan, M.J., Grove, D., Koblin, B.A., Buchbinder, S.P., Keefer, M.C., Tomaras, G.D., et al.; HVTN 505 Study Team (2013). *N. Engl. J. Med.* 369, 2083–2092.
- Hansen, S.G., Piatak, M., Jr., Ventura, A.B., Hughes, C.M., Gilbride, R.M., Ford, J.C., Oswald, K., Shoemaker, R., Li, Y., Lewis, M.S., et al. (2013). *Nature* 502, 100–104.
- Haynes, B.F., Gilbert, P.B., McElrath, M.J., Zolla-Pazner, S., Tomaras, G.D., Alam, S.M., Evans, D.T., Montefiori, D.C., Karnasuta, C., Sutthent, R., et al. (2012). *N. Engl. J. Med.* 366, 1275–1286.
- Letvin, N.L., Rao, S.S., Montefiori, D.C., Seaman, M.S., Sun, Y., Lim, S.Y., Yeh, W.W., Asmal, M., Gelman, R.S., Shen, L., et al. (2011). *Sci. Transl. Med.* 3, 81ra36.
- Pitisuttithum, P., Gilbert, P., Gurwith, M., Heyward, W., Martin, M., van Griensven, F., Hu, D., Tappero, J.W., and Choopanya, K.; Bangkok Vaccine Evaluation Group (2006). *J. Infect. Dis.* 194, 1661–1671.
- Rerks-Ngarm, S., Pitisuttithum, P., Nitayaphan, S., Kaewkungwal, J., Chiu, J., Paris, R., Premsri, N., Namwat, C., de Souza, M., Adams, E., et al.; MOPH-TAVEG Investigators (2009). *N. Engl. J. Med.* 361, 2209–2220.
- Reynolds, M.R., Weiler, A.M., Piaskowski, S.M., Piatak, M., Jr., Robertson, H.T., Allison, D.B., Bett, A.J., Casimiro, D.R., Shiver, J.W., Wilson, N.A., et al. (2012). *Vaccine* 30, 4465–4475.
- Rolland, M., Edlefsen, P.T., Larsen, B.B., Tovana-butra, S., Sanders-Buell, E., Hertz, T., deCamp, A.C., Carrico, C., Menis, S., Magaret, C.A., et al. (2012). *Nature* 490, 417–420.

Genetic Variation Meets Replication Origins

Ronald J. Hause¹ and Jay Shendure^{1,*}

¹Department of Genome Sciences, University of Washington, Seattle, WA 98195, USA

*Correspondence: shendure@uw.edu

<http://dx.doi.org/10.1016/j.cell.2014.11.009>

Genome replication programs are highly orchestrated. In this issue, Koren and colleagues leverage whole-genome sequencing data to discover that DNA replication timing patterns differ between individuals and are associated with genetic variants. These so-called rtQTLs represent a new form of quantitative trait loci that may influence gene dosage and mutational frequency and provide mechanistic insights into the regulation of DNA replication.

DNA replication, the process of producing two identical replicas from one DNA template, is the basis for biological inheritance in all living organisms. Over 60 years ago, early techniques of radiolabeling were used to determine that most DNA synthesis occurred in a restricted period of the cell cycle, now termed the “S” phase for “synthesis.” The observation that the inactive X chromosome in female placental mammals was late replicating and corresponded to the hypercondensed “Barr body” indicated that the timing of DNA replication is locus-specific and functionally important. Replication origin usage also varies across cell differentiation programs and is strongly correlated with three-dimensional chromosomal architecture within the nucleus (Ryba et al., 2010). Notably, the replication timing program is also highly unstable in human cancers and associated with increased translocation and somatic mutation rates (Méchalí et al., 2013). However, we still know relatively little regarding the specific mechanisms orchestrating the replication timing program across the genomes of higher eukaryotes.

It has recently been shown that patterns of replication timing can be inferred from comparing the ratios of read depths obtained from S phase cells to G1 phase cells, with sites of earlier replication exhibiting greater sequence coverage in S phase cells (Van der Aa et al., 2013; Koren et al., 2012). While global replication timing profiles were highly correlated across six individual cell lines they previously examined (Koren et al., 2012),

several questions remained: To what extent, if at all, do patterns in replication timing differ between humans? Are there genetic determinants of differences in DNA replication timing? If so, what are the broader implications of variation in replication timing for other aspects of genome regulation?

In this issue of *Cell*, Koren et al. take a large step forward from their earlier work, leveraging whole-genome sequencing data from 161 individuals sequenced by the 1000 Genomes Project to examine variation in DNA replication timing in humans (Koren et al., 2014). They demonstrate that variation in sequencing coverage across chromosomes from proliferating lymphoblastoid cells is correlated with previous estimates of replication timing and thus reflects active DNA replication (Figure 1A). Koren et al. then identified and, as importantly, validated by replication, 16 *cis* replication timing quantitative trait loci (rtQTLs) associated with either the loss/gain of a replication origin (Figure 1B), variable usage or activation timing of a replication origin (Figure 1C), or variable replication initiation zone lengths (Figure 1D). Remarkably, these ~20 kb haplotypes are associated with replication timing variation up to 2 Mb away. These rtQTLs are also correlated with the expression levels of transcripts up to hundreds of kilobases away, with individuals with earlier DNA replication having higher mRNA expression levels nearby. Lastly, the authors identify and discuss a notable association between alleles that predispose to *JAK2* mutations, commonly implicated in

hematological malignancies, and pronounced replication. Early replication at this site occurs in the opposite direction of *JAK2* transcription, suggesting that its fragility might be enhanced by collisions between the replication and transcription machinery.

The authors also identified 17 putative *trans* associations but discarded them because they were likely artifactual associations due to population structure. Future work incorporating larger sample sizes and correcting for population structure using principal component analysis or linear mixed models may allow us to better assess whether *trans*-acting polymorphic genetic factors are involved in coordinating replication timing.

The novel approach of Koren et al. (2014) of inferring replication timing and origins directly from whole-genome sequence coverage of proliferating cells offers researchers the opportunity to investigate many unanswered questions without the need for isolating cells in specific stages or labeling newly synthesized nucleotides. To the extent that the corresponding genome sequences derive from proliferating cells, this approach may be readily extendable to shotgun genome sequences currently available, such as The Cancer Genome Atlas (TCGA), the Encyclopedia of DNA Elements (ENCODE), and The Genotype-Tissue Expression project (GTEx), to examine the relevance of DNA replication timing variation in cancer, functional regulatory elements throughout the genome, and gene regulation across tissues. For instance, previous work has

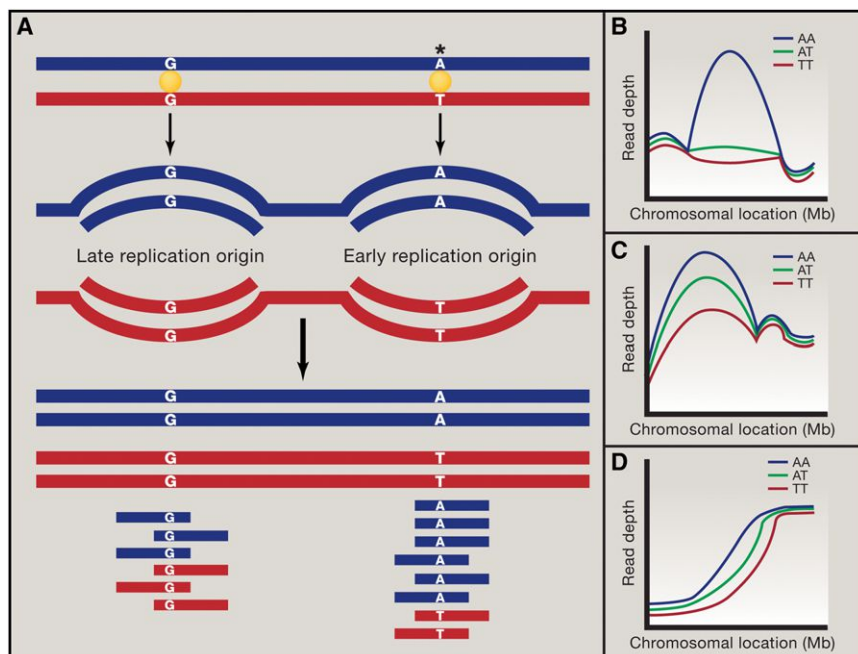


Figure 1. Inferring Variation in Replication Timing from Whole-Genome Sequencing

(A) DNA replication is initiated at replication origins (yellow circles). Early replication origins (right) synthesize more DNA than late replication origins (left) that can be observed by different sequencing read depths. Replication timing quantitative trait loci (rtQTLs) can result in allele-specific variation in DNA replication, indicated as more reads derived from the "A" than the "T" haplotype from the early replication origin.

(B) Approximately half of variant regions differed by the presence or absence of a peak, indicating gain or loss of a replication origin or origin cluster.

(C) ~25% of variant regions involved differences in peak height, indicating variable usage or activation timing of a replication origin.

(D) The remaining quarter involved a shift in the replication slope region, indicating variable replication initiation zone lengths.

"AA," "AT," and "TT" designate arbitrary homozygous alleles and their heterozygous intermediates associated as rtQTLs at three different variable regions in (B–D).

demonstrated that replication origin activity is associated with open chromatin structure and correlated globally with gene expression across a genome (Hiratani et al., 2008). Koren et al. observed enrichment of rtQTLs within enhancers active in lymphoblastoid cell lines and mRNA expression QTLs (eQTLs) and speculate that these rtQTLs could alter chromatin structure or DNA sequences bound by factors involved in replication initiation. As such, next steps should include examining the relationships between rtQTLs and DNaseI hypersensitivity QTLs, histone modification QTLs, and nucleosome positioning as potential mechanisms underlying these *cis*-acting associations. Furthermore, as with other

genetic determinants of molecular functionality, it will be interesting to explore the extent to which rtQTLs are global, tissue-specific, population-specific, or temporal at different stages of development.

Koren et al. here identify a striking relationship between shotgun genome sequence coverage and replication timing and exploit this relationship to define a new type of QTL. These rtQTLs can potentially elucidate mechanisms connecting common variants associated with complex phenotypes, such as human disease and drug response, with intermediate phenotypes, such as mRNA expression levels and chromatin accessibility states. Given the previously known relationship between replication timing

and increased mutational rates (Koren et al., 2012; Stamatoyannopoulos et al., 2009), rtQTLs could also affect the probability of somatic mutations at specific loci involved in human disease and cancer susceptibility, such as in the authors' *JAK2* example; therefore, quantifying DNA replication timing in association with mutation rate could address what proportion of mutation load is because of differential replication origin usage or efficiency. More generally, future work examining the factors involved in regulating replication timing could allow us to better understand the causes and consequences of mutation rate variation throughout the genome. Lastly, in addition to facilitating the identification of the causal variants underlying rtQTLs and other molecular QTLs, genome editing approaches such as CRISPR/Cas9 may allow us to examine the molecular consequences of directly manipulating replication timing in proliferating cells.

ACKNOWLEDGMENTS

R.J.H. and J.S. are supported by the National Institutes of Health (DP1HG007811).

REFERENCES

- Hiratani, I., Ryba, T., Itoh, M., Yokochi, T., Schwaiger, M., Chang, C.-W., Lyuu, Y., Townes, T.M., Schübeler, D., and Gilbert, D.M. (2008). *PLoS Biol.* 6, e245.
- Koren, A., Polak, P., Nemesh, J., Michaelson, J.J., Sebat, J., Sunyaev, S.R., and McCarroll, S.A. (2012). *Am. J. Hum. Genet.* 91, 1033–1040.
- Koren, A., Handsaker, R.E., Kamitaki, N., Karlic, R., Ghosh, S., Polak, P., Eggan, K., and McCarroll, S.A. (2014). *Cell* 159, this issue, 1015–1026.
- Méchali, M., Yoshida, K., Coulombe, P., and Pasero, P. (2013). *Curr. Opin. Genet. Dev.* 23, 124–131.
- Ryba, T., Hiratani, I., Lu, J., Itoh, M., Kulik, M., Zhang, J., Schulz, T.C., Robins, A.J., Dalton, S., and Gilbert, D.M. (2010). *Genome Res.* 20, 761–770.
- Stamatoyannopoulos, J.A., Adzhubei, I., Thurman, R.E., Kryukov, G.V., Mirkin, S.M., and Sunyaev, S.R. (2009). *Nat. Genet.* 41, 393–395.
- Van der Aa, N., Cheng, J., Mateiu, L., Zamani Esteki, M., Kumar, P., Dimitriadou, E., Vanneste, E., Moreau, Y., Vermeesch, J.R., and Voet, T. (2013). *Nucleic Acids Res.* 41, e66.

HIV-1 Packing to Leave

Bryan R. Cullen^{1,*}

¹Department of Molecular Genetics and Microbiology and Center for Virology, Duke University Medical Center, Durham, NC 27710, USA

*Correspondence: bryan.cullen@duke.edu

<http://dx.doi.org/10.1016/j.cell.2014.11.008>

HIV-1 virion assembly at the plasma membrane requires the selective recruitment of the viral RNA genome into nascent viral particles while cellular transcripts are excluded. Kutluay et al. now demonstrate that this is a two-step process in which Gag binds sequentially to different sites on the viral genome.

The final step in the life cycle of HIV-1 is the assembly and release of progeny virions by budding through the infected cell plasma membrane (Sundquist and Kräusslich, 2012). In this issue of *Cell*, Kutluay et al. (2014) demonstrate that the selective recruitment of the viral RNA genome into virions—mediated by the nucleocapsid (NC) domain of the viral Gag polyprotein—involves the sequential interaction of Gag with distinct genomic RNA target sites, first in the infected cell cytoplasm and then during the formation of nascent viral particles at the plasma membrane. Moreover, they also show that transient binding of tRNAs to the matrix (MA) domain of Gag plays a key role in targeting Gag to the plasma membrane.

Analysis of the requirements for selective recruitment of an RNA into HIV-1 virions has previously identified a structured RNA element called “psi” (“ ψ ”), located in the viral RNA 5' UTR, that is selectively bound by NC and that greatly facilitates genomic RNA packaging (Clever et al., 1995; Lu et al., 2011). However, the specificity of the ψ :NC interaction is modest, and regions outside of ψ have been proposed to also facilitate genomic RNA packaging (Clever et al., 1995; Richardson et al., 1993). Moreover, while the initial interaction of Gag monomers with the genomic RNA occurs in the cytoplasm, the formation of multimeric Gag complexes on the RNA genome is dependent on localization to the plasma membrane (Kutluay and Bieniasz, 2010).

To address which RNA:protein interactions are involved in HIV-1 virion morphogenesis and whether these are modulated in different cellular compartments, i.e., in the cytoplasm, at the plasma membrane, and inside HIV-1 virions, Kutluay et al.

(2014) developed a remarkably efficient variation on the previously published “photoreactive ribonucleoside-enhanced crosslinking and immunoprecipitation” (PAR-CLIP) technique (Hafner et al., 2010). PAR-CLIP relies on the incorporation of the photoreactive uridine analog 4-thiouridine into newly synthesized RNA molecules, which confers efficient crosslinking upon brief UV irradiation. Immunoprecipitation, using antibodies specific for Gag or for individual Gag domains, combined with RNase treatment and deep sequencing, then allowed a comprehensive analysis of the RNA sequences bound by these proteins.

Initial analysis gratifyingly reveals strong binding of Gag to the ψ sequence but also demonstrates strong binding to the highly structured Rev response element (RRE), which is required for HIV-1 genomic RNA export to the cytoplasm and serves as the binding site for the HIV-1 Rev protein (Malim et al., 1989). However, only ~5% of RNA binding by Gag involves viral transcripts and the enrichment of viral RNA in the Gag-bound versus total RNA pool is only 3- to 5-fold relative to cellular RNAs, consistent with the relatively low in vitro specificity of Gag for ψ (Richardson et al., 1993). In contrast, analysis of HIV-1 Gag binding to RNA in immature virions gives a very different picture. In particular, virion RNA now contributed ~50% of all Gag binding sites and these were scattered across the entire viral RNA genome. These binding sites are not, however, random and in fact reveal a clear preference for “A”-rich RNA sequences. Together, these data reveal a two-step process for virion RNA recruitment into virions. In the cytoplasm, largely monomeric Gag binds to both cellular RNAs and,

with modest specificity, to viral genomic RNAs. Both cellular and viral RNA-bound Gag molecules are then recruited to the inner surface of the plasma membrane (see below), where they have the potential to serve as nucleation sites for Gag multimerization. However, this process requires the specific interaction of Gag with A-rich sequences, scattered across the viral genome, that are absent on most cellular transcripts, thus explaining the selectivity of genomic RNA packaging.

While the NC domain of Gag is required for genomic RNA recruitment into virions, the amino-terminal MA domain is responsible for the recruitment of Gag, and any NC-bound RNA species, to the plasma membrane (Chukkapalli et al., 2010). Gag recruitment to the plasma membrane requires the interaction of a highly basic region (HBR) in MA with acidic phospholipids, especially phosphatidylinositol 4, 5-bisphosphate [PtdIns(4,5)P₂], that are only present on the plasma membrane. The HBR of MA has previously been reported to also bind RNA nonspecifically, and RNA had been proposed to compete with membrane binding by Gag, thus favoring Gag recruitment to high-affinity binding sites on the plasma membrane and blocking binding to intracytoplasmic membranes (Chukkapalli et al., 2013; Chukkapalli et al., 2010). The data presented by Kutluay et al. (2014) extend these earlier in vitro observations by showing that MA indeed binds RNA efficiently in infected cells. More importantly, they identified the bound RNAs as almost exclusively tRNAs. The tRNAs bound to MA are, however, released during virion formation, and it is therefore unlikely that MA facilitates the virion incorporation of the

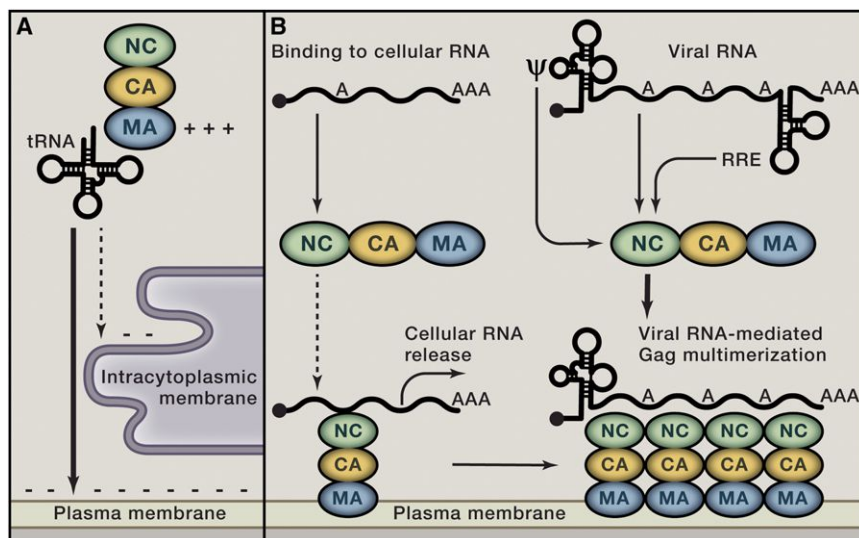


Figure 1. RNA Binding to the HIV-1 Gag Protein

(A and B) Schematic representation of how tRNA binding to the Gag matrix (MA) domain facilitates recruitment of Gag to the plasma membrane (A) and the two step process by which the nucleocapsid (NC) domain of Gag recruits HIV-1 genomic RNAs into newly formed virions (B). See text for detailed discussion. CA, capsid domain of Gag; ψ , psi RNA packaging signal; RRE, Rev response element.

lysine tRNA primer, especially as this tRNA is not selectively bound by MA.

In conclusion, this report (Kutluay et al., 2014) pulls together a number of previous in vitro studies and substantially extends these earlier data using highly sensitive analytical techniques in HIV-1 infected cells. These data reveal that two RNA binding domains in Gag, located in MA and NC, both play a key role in virion formation (Figure 1). tRNA binding by MA blocks Gag recruitment to intracellular membranes and instead promotes Gag binding to the negatively charged plasma membrane. Simultaneously, the

NC domain initially recruits a mixture of cellular and, preferentially, viral transcripts to the plasma membrane, where viral genomic RNAs are then incorporated into virions due to selective binding of NC to specific A-rich sequences on the viral RNA genome during Gag multimerization (Figure 1).

One question left unresolved is whether viral RNA binding sites for Gag other than ψ , for example in the RRE, actually promote the recruitment of cytoplasmic RNA into virions, as also suggested by earlier work (Richardson et al., 1993). Kutluay et al. (2014) do show that deletion

of the RRE, combined with substitution by an unrelated nuclear RNA export signal to allow the HIV-1 genomic RNA to reach the cytoplasm, fails to uncover any role for the RRE in viral RNA packaging when the ψ element is intact. Nevertheless, it remains possible that the RRE, or other regions in the HIV-1 genome, are largely redundant with ψ so that a role for these RNA sequences in viral RNA packaging would only be uncovered when ψ is partially or fully inactivated.

REFERENCES

- Chukkapalli, V., Oh, S.J., and Ono, A. (2010). *Proc. Natl. Acad. Sci. USA* 107, 1600–1605.
- Chukkapalli, V., Inlora, J., Todd, G.C., and Ono, A. (2013). *J. Virol.* 87, 7155–7159.
- Clever, J., Sasseti, C., and Parslow, T.G. (1995). *J. Virol.* 69, 2101–2109.
- Hafner, M., Landthaler, M., Burger, L., Khorshid, M., Hausser, J., Berninger, P., Rothballer, A., Ascano, M., Jr., Jungkamp, A.C., Munschauer, M., et al. (2010). *Cell* 141, 129–141.
- Kutluay, S.B., and Bieniasz, P.D. (2010). *PLoS Pathog.* 6, e1001200.
- Kutluay, S.B., Zang, T., Blanco-Melo, D., Powell, C., Jannain, D., Errando, M., and Bieniasz, P.D. (2014). *Cell* 159, this issue, 1096–1109.
- Lu, K., Heng, X., Garyu, L., Monti, S., Garcia, E.L., Kharytonchyk, S., Dorjsuren, B., Kulandaivel, G., Jones, S., Hiremath, A., et al. (2011). *Science* 334, 242–245.
- Malim, M.H., Hauber, J., Le, S.Y., Maizel, J.V., and Cullen, B.R. (1989). *Nature* 338, 254–257.
- Richardson, J.H., Child, L.A., and Lever, A.M. (1993). *J. Virol.* 67, 3997–4005.
- Sundquist, W.I., and Kräusslich, H.G. (2012). *Cold Spring Harb. Perspect. Med.* 2, a006924.

A Molecular Gauge for Nitrogen Economy in Plants

Timothy R. Fallon^{1,2} and Jing-Ke Weng^{1,2,*}

¹Whitehead Institute for Biomedical Research, Nine Cambridge Center, Cambridge, MA 02142, USA

²Department of Biology, Massachusetts Institute of Technology, Cambridge, MA 02139, USA

*Correspondence: wengj@wi.mit.edu

<http://dx.doi.org/10.1016/j.cell.2014.11.007>

Understanding nitrogen metabolism in plants holds promise for future agricultural improvements. Chellamuthu et al. now identify a feedback regulation in plant nitrogen metabolism through glutamine sensing. This mechanism appears to be conserved from algae to flowering plants with a few surprising exceptions.

At the turn of the 19th century, newly elected president of the British Academy of Science William Crookes gave a dire prediction to the assembled academy. He stated bluntly, "...all civilized nations stand in deadly peril of not having enough to eat" (Crookes, 1898). Crookes argued that a lack of fertilizer, in particular fixed nitrogen, would fatally limit food production and bring about global starvation. Thankfully, the development of the Haber-Bosch process in the early decades of the 20th century allowed for the catalyzed fixation of atmospheric nitrogen into active species suitable for use in fertilizer at an industrial scale. Combined with the advances in plant breeding and agricultural technology from the "Green Revolution" of the mid-20th century, we continue today to stay ahead of Crookes' dire predictions. As the human population continues to expand, further understanding and optimization of plant nitrogen utilization remains a pressing task. Toward this end, Chellamuthu et al. (2014) in this issue of *Cell* describe a significant advance in the understanding of the molecular feedback control of nitrogen metabolism across the plant kingdom.

Beyond using nitrogen for protein or nucleotide synthesis alone, many plants also produce large quantities of nitrogen containing specialized metabolites known as alkaloids, caffeine being a well-known example. Plants have also evolved to store excess nitrogen in both seeds and vegetative tissues in the form of specialized storage proteins. These proteins store nitrogen both in the amide backbone of the polypeptide, as well as by having a high proportion of amino acids

enriched with nitrogen containing R groups, such as arginine, asparagine, and glutamine (Staswick, 1994). Unlike vertebrate nitrogen metabolism, which, particularly in the case of carnivores, often operates in a state of nitrogen excess tuned toward excretion, plants, algae, and bacteria must assimilate inorganic nitrogen from the environment and therefore have to carefully balance nitrogen utilization. Perhaps unsurprisingly, studies of nitrogen metabolism in these organisms have revealed elegant regulatory mechanisms.

In the case of bacteria, previous studies indicated that nitrogen metabolism was regulated through a key conserved protein, known as P_{II} (Forchhammer, 2007). P_{II} integrates concentration information of key metabolites, including ATP/ADP, 2-oxoglutarate, and glutamine, in order to regulate downstream proteins in nitrogen metabolism. Providing a chemical logic for this feedback, 2-oxoglutarate, ATP, and glutamine all participate in the dominant metabolic pathway for nitrogen assimilation, the glutamine synthase/glutamine oxoglutarate aminotransferase (GS/GOGAT) pathway (Forchhammer, 2007). While P_{II} senses ATP and 2-oxoglutarate levels through a direct binding interaction, glutamine sensing is transduced through a phylogenetically divergent array of secondary sensors, which posttranslationally modify P_{II} in response to glutamine levels. P_{II} then binds and regulates downstream enzymes in nitrogen metabolism, most notably the committed step in arginine biosynthesis, N-acetylglutamate kinase (NAGK) (Heinrich et al., 2004). NAGK is feedback inhibited by

arginine, its ultimate downstream product. P_{II} binding activates NAGK by relieving this inhibitory feedback of arginine.

The chloroplasts of plants share ancestry with modern cyanobacteria, and plant P_{II} has been conserved throughout the long evolutionary divergence between plants and bacteria (Chellamuthu et al., 2013). Genetic studies in the model flowering plant *Arabidopsis thaliana* indicated that loss of P_{II} produced a number of nitrogen metabolism-related phenotypes (Ferrario-Méry et al., 2006). Furthermore, in vitro biochemical and structural analyses demonstrated remarkable similarity of P_{II} with its bacterial counterparts, including the interaction with NAGK (Mizuno et al., 2007). However, the question remained if the divergent mechanisms of secondary sensing of glutamine through posttranslational modifications of P_{II} were conserved.

To better define the regulatory role of P_{II} in plant nitrogen metabolism, Chellamuthu et al. (2014) now undertake a biochemical study on NAGKs and their corresponding P_{II} partners from both *A. thaliana* and the model green algae *Chlamydomonas reinhardtii*. Like the P_{II}-NAGK interaction in bacteria, previous studies had shown that *A. thaliana* P_{II} constitutively relieves arginine feedback inhibition of NAGK (Beez et al., 2009). However, the experiments of Chellamuthu et al. with *C. reinhardtii* P_{II} and NAGK produce a surprising finding. In contrast to *A. thaliana* and bacterial P_{II}, *C. reinhardtii* P_{II} does not relieve arginine feedback inhibition of NAGK. Chellamuthu et al. surmise there might be a missing small molecule in their assays that P_{II}

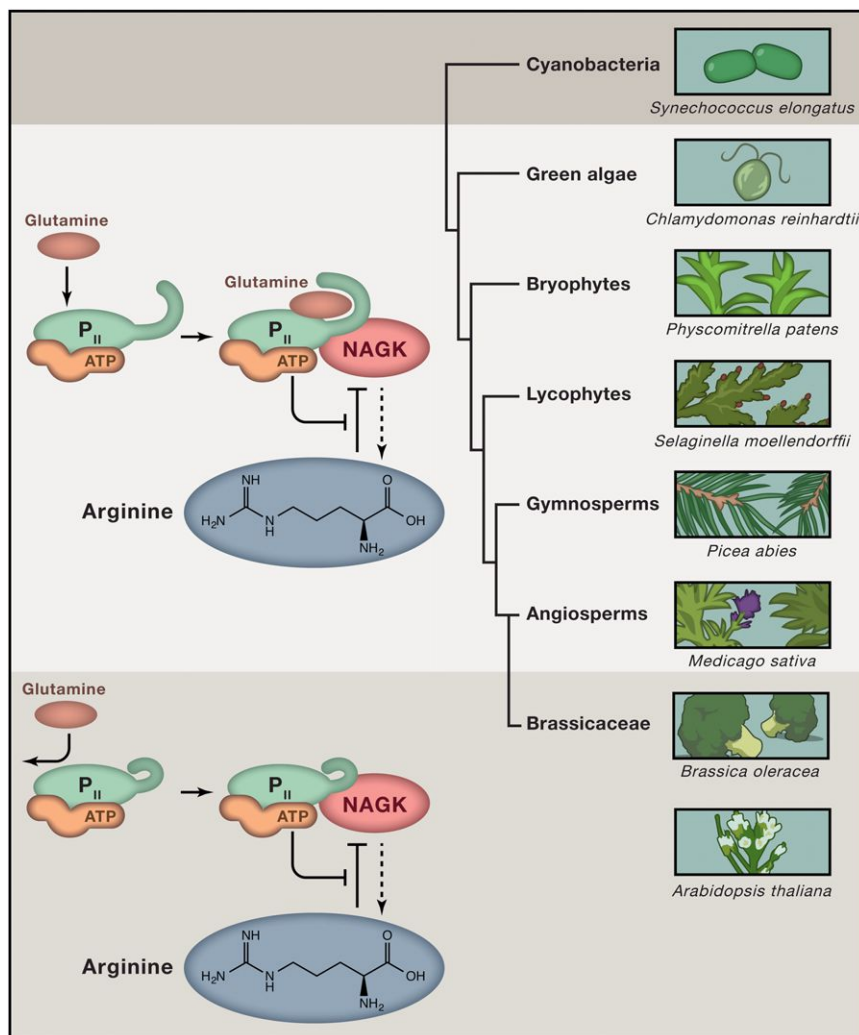


Figure 1. P_{II} Proteins Directly Sense Glutamine for NAGK Activation and Arginine Biosynthesis in Plants, Except in Brassicaceae

P_{II} activates arginine biosynthesis by antagonizing the feedback inhibition of arginine upon the committed step of arginine biosynthesis, NAGK. In every plant taxa except *Brassicaceae*, the family that contains the model flowering plant *Arabidopsis thaliana*, glutamine is directly sensed by binding to a conserved C-terminal extension, termed the Q loop, of P_{II}. In the case of *Brassicaceae*, glutamine is not required for activation of NAGK by P_{II}. In both cases, bound ATP appears to be required for P_{II} to antagonize NAGK feedback inhibition.

could be sensing. After assaying a number of primary metabolites, they identify that addition of glutamine rescues the *C. reinhardtii* P_{II}-NAGK interaction. Glutamine, the metabolite previously known only to be sensed by P_{II} through secondary proteins, appears to be sensed directly, revealing a dramatic mechanistic divergence in P_{II} between plants and bacteria.

Curiously, plant P_{II} proteins have a functionally uncharacterized but conserved C-terminal extension, which is absent in

bacteria. Through X-ray crystallography, it is clear that the C-terminal extension in *C. reinhardtii* P_{II} forms a so-called Q loop region, which constitutes a large proportion of the glutamine-binding site. Interestingly the Q loop of *A. thaliana* and other *Brassicaceae* contains a three-amino-acid deletion, rendering their P_{II} unable to directly sense glutamine. In *C. reinhardtii*, NAGK activation through P_{II} binding is glutamine-dependent, whereas deletion of the Q loop region in P_{II} completely abolishes its ability to bind to NAGK. Measured

in vitro, the EC₅₀ of glutamine for NAGK activation by P_{II} is in the millimolar range, which, as noted by the authors, is comparable to reported intracellular concentrations of glutamine. Future research is required to demonstrate the physiological relevance of this sensing mechanism in vivo.

Based on phylogenetic sequence analysis and in vitro studies of representative members of diverse lineages, the P_{II} Q loop and direct glutamine sensing appears to be widely conserved in the plant kingdom (Figure 1). However, many interesting questions arise when considering the functional relevance of the unexpected divergence of P_{II} within *Brassicaceae*. For instance, are the observed Q loop deletions an adaptive change or a deleterious mutation that became fixed in the common ancestor of *Brassicaceae*? Could *Brassicaceae* contain alternative glutamine-sensing mechanisms to compensate for the loss of glutamine-sensing capability in its P_{II}? In vivo studies of these phenomena would be important to reveal if P_{II} divergence represents an exploitable trait for genetic engineering of crops.

ACKNOWLEDGMENTS

J.K.W. is a Pew Scholar in the Biomedical Sciences.

REFERENCES

- Beez, S., Fokina, O., Herrmann, C., and Forchhammer, K. (2009). *J. Mol. Biol.* 389, 748–758.
- Chellamuthu, V.R., Alva, V., and Forchhammer, K. (2013). *Planta* 237, 451–462.
- Chellamuthu, V.-R., Ermilova, E., Lapina, T., Lüddecke, J., Minaeva, E., Herrmann, C., Hartmann, M.D., and Forchhammer, K. (2014). *Cell* 159, this issue, 1188–1199.
- Crookes, W. (1898). *Science* 8, 561–575.
- Ferrario-Méry, S., Besin, E., Pichon, O., Meyer, C., and Hodges, M. (2006). *FEBS Lett.* 580, 2015–2020.
- Forchhammer, K. (2007). *Front. Biosci.* 12, 358–370.
- Heinrich, A., Maheswaran, M., Ruppert, U., and Forchhammer, K. (2004). *Mol. Microbiol.* 52, 1303–1314.
- Mizuno, Y., Moorhead, G.B.G., and Ng, K.K.-S. (2007). *J. Biol. Chem.* 282, 35733–35740.
- Staswick, P.E. (1994). *Annu. Rev. Plant Biol.* 45, 303–322.

The Ribosome Emerges from a Black Box

V. Ramakrishnan^{1,*}

¹MRC Laboratory of Molecular Biology, Cambridge CB2 0QH, UK

*Correspondence: ramak@lmb.cam.ac.uk

<http://dx.doi.org/10.1016/j.cell.2014.10.052>

Although the basic facts about the ribosome were already known 40 years ago, elucidating its atomic structure and molecular mechanisms required sheer persistence and the innovative use of new technology and methods. These advances have transformed our understanding of translation in the cell.

Two years after *Cell* was born, I left theoretical physics to become a graduate student in biology at UCSD, and two years after that, I joined Peter Moore's laboratory at Yale to begin working on the ribosome. So by a strange coincidence, my life as a ribosome scientist has largely overlapped with the life of *Cell*. It is interesting to look back at what was known about the ribosome at *Cell*'s inception, and the remarkable progress that has been made since.

That progress can be visually encapsulated in Figure 1, which shows the structure of the ribosome 40 years ago (Lake, 1976) compared to a recent atomic structure of a translating ribosome bound to the translocon (Voorhees et al., 2014). The latter was obtained by single-particle electron cryomicroscopy (cryoEM), a technique that not only did not exist in 1974, but which only recently has been capable of producing high-resolution structures of asymmetric objects like the ribosome.

The state of the ribosome in 1974 was captured in a comprehensive book, entitled simply "Ribosomes," with contributions by many leading scientists of the day (Nomura et al., 1974). It marked the end of what is commonly referred to as the "golden age" of molecular biology, during which the salient facts about the ribosome had already been established. Ribosomes from all species consist of consist of two subunits (30S and 50S for bacteria and 40S and 60S for eukaryotes) and in most species are two-thirds of RNA by mass. They contain over 50 proteins, and three large ribosomal RNA (rRNA) molecules. The small subunit binds mRNA and the large subunit carries out peptidyl transfer. The tRNAs bound in the P and A sites hold the nascent peptide chain and the new amino acid to be added, respectively. Protein factors act at each stage of the process, and many of these are GTPases. These basic facts have been textbook material ever since.

Given the complexity of the ribosome, to go beyond that basic level of understanding seemed a daunting task, so only a few aficionados persisted in working on it. During the subsequent decades, three qualitative advances have changed our understanding of the ribosome. The first was the emergence of the idea that the ribosome is primarily an RNA-based machine. The second was the atomic structure of the ribosome, which not only confirmed that idea, but paved the way for ever more sophisticated experiments to understand the mechanism of translation. Finally, we are now beginning to understand the ribo-

some as a dynamic machine in which large conformational changes are essential to its function.

The Ribosome as an RNA Machine

Early on, it was thought the many different proteins of the ribosome might be responsible for its various functions, with the RNA as a scaffold to hold the various proteins in place. However, the ribosome poses the classic "chicken or egg" question: If the ribosome consists of both RNA and protein, and is needed to make protein, how did it originate? Crick (1968) rather presciently wrote, "It is tempting to wonder if the primitive ribosome could have been made *entirely* of RNA" (original italics). To my knowledge, this was the first idea that RNA could both carry genetic information and perform catalysis and can be thought of as the origin of the "RNA world hypothesis," which postulates a primordial world consisting of replicating RNA molecules before the advent of proteins.

The first indication that rRNA contributes to translation at all arose from the discovery that complementarity between the 3' end of 16S rRNA and a sequence on mRNA upstream of the start codon was important for proper initiation of translation (Steitz, 1969; Shine and Dalgarno, 1974). However, base pairing was already known to be a feature of nucleic acids and unrelated to the broader functions of the ribosome such as catalysis, ligand binding, and movement. Similarly, modification of rRNA, but not protein, was shown to affect tRNA binding, but in the absence of any evidence for catalytic properties of rRNA, it was suggested that the binding sites must consist of both RNA and protein (Noller and Chaires, 1972).

The sequencing of 16S and 23S ribosomal RNAs (Brosius et al., 1978; Brosius et al., 1980) enabled the establishment of their secondary structure (Woese et al., 1980; Glotz and Brimacombe, 1980) and had consequences for biology far beyond the ribosome: they paved the way for the discovery of a third branch of life, the archaea (Woese et al., 1990).

In conjunction with the ability to reconstitute the ribosome from purified components (Held et al., 1973; Nierhaus and Dohme, 1974), knowledge of the sequence of rRNA also laid the groundwork for the use of chemical footprinting methods to study its interaction with various ligands (Moazed and Noller, 1986; Moazed and Noller, 1989a). Such studies, as well as complementary crosslinking methods (Döring et al., 1994), identified regions of rRNA that interact with tRNAs in the A, P, and E sites.

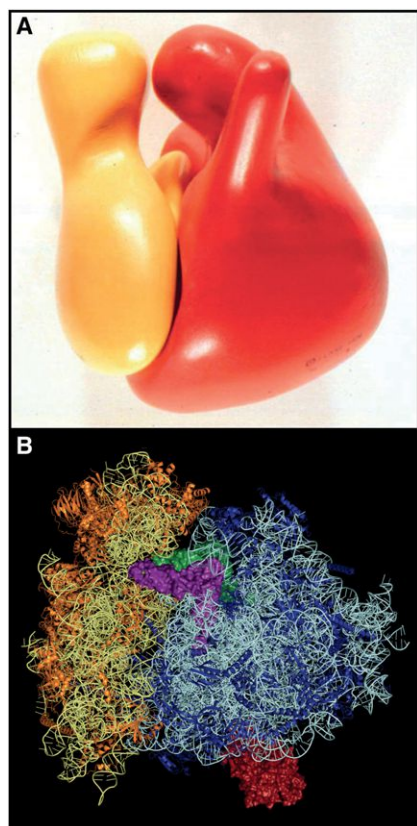


Figure 1. The Ribosome Then and Now

(A) Model of the *E. coli* ribosome 40 years ago, obtained by negative stain electron microscopy (Lake, 1976). The small subunit is shown in yellow and the large subunit in red.

(B) High-resolution structure by single-particle cryoEM of a translating mammalian ribosome (small subunit in yellow and large in blue) bound to the translocon (red) and A- and P-site tRNAs in purple and green respectively (Voorhees et al., 2014).

These studies provided initial physical constraints for building models of the ribosome.

Further evidence for functional roles for rRNAs came from studies on antibiotic binding sites in the ribosome. The first antibiotic resistance mutation mapped (to streptomycin) was on ribosomal protein S12 (Traub and Nomura, 1968), but in hindsight this was misleading, since antibiotics have little affinity for ribosomal proteins. Subsequent chemical footprinting showed that antibiotics, including streptomycin, interact with specific and distinct sites on ribosomal RNA (Moazed and Noller, 1987a; Moazed and Noller, 1987b).

Together, these studies suggested that rRNA was likely to be of functional importance. However, the notion of the ribosome as fundamentally an RNA enzyme really gained plausibility only when RNA catalysis was discovered in the context of the group I intron (Zaug and Cech, 1986) and RNase P (Guerrier-Takada et al., 1983). These discoveries led to a resurgence of interest in the ribosome and ribosomal RNA in particular (Moore, 1988).

Unlike the case with the simpler group I intron and RNase P, proving that catalysis in the ribosome was a property of RNA proved difficult. When thermophilic 50S subunits were treated

extensively with protease and phenol extraction to remove proteins, they nevertheless retained peptidyl transferase activity, strongly suggesting a catalytic role for RNA (Noller et al., 1992). However, even after extensive protease treatment, several peptide fragments and even some entire proteins resisted digestion and extraction, and the complete removal of proteins resulted in a loss of activity. So, as implicitly acknowledged by the authors in their cautiously worded title, “Unusual resistance of peptidyl transferase to protein extraction procedures,” these experiments were not conclusive. The limitations of such experiments were foreseen by Crick, who, when he suggested that primitive ribosomes may have consisted entirely of RNA, also said, “Without a detailed knowledge of the structure of present-day ribosomes it is difficult to make an informed guess” (Crick, 1968). As discussed below, that structure took many decades after the discovery of the ribosome.

Toward the Atomic Structure of the Ribosome

Forty years ago, the gross morphology of the two ribosomal subunits had been determined by conventional negative-stain electron microscopy with the 30S subunit having a “head,” a “platform,” and a “body” and the large subunit having a central protuberance flanked by two stalks. But strikingly, even essential features such as the number and location of all of the tRNA binding sites and the path of the nascent peptide were established only gradually.

tRNA Binding Sites

Aminoacyl tRNAs are the substrates of the ribosome, responsible for delivering each new amino acid to the growing polypeptide chain. Early on, it was accepted that the ribosome would have two tRNA binding sites, a P site occupied by a tRNA holding the nascent peptide and an A site for the tRNA delivering the new amino acid corresponding to the codon on mRNA. A third site, called the E (exit) site, into which the deacylated tRNA moves prior to being ejected from the ribosome, was first proposed a long time ago (Wettstein and Noll, 1965), but even its existence remained controversial until convincingly demonstrated by Nierhaus and coworkers (Rheinberger et al., 1981). The fact that the E site spans both the small and large subunit became universally accepted only after more detailed structures of the ribosome by cryoEM began to emerge (Agrawal et al., 1996; Stark et al., 1997). These studies also showed that the mRNA snakes around a cleft in the small subunit.

Passage of the Nascent Peptide through a Tunnel in the Large Subunit

Early cartoons of the ribosome and a more recent sculpture at Cold Spring Harbor Laboratory simply show the peptide chain being extruded from the intersubunit space. The existence of a tunnel in the large subunit through which the nascent peptide must emerge was inferred by localizing antibodies to the nascent peptide to a region opposite the intersubunit interface (Bernabeu and Lake, 1982). More direct evidence for an exit tunnel came from electron microscopy of two-dimensional crystalline sheets or sections (Milligan and Unwin, 1986; Yonath et al., 1987), but the tunnel was firmly established only when cryoEM reconstructions of sufficient resolution became available (Frank et al., 1995), and its atomic nature was elucidated by analysis of the crystal structure of the large subunit (Ban et al., 2000; Nissen

et al., 2000). The discovery of the exit tunnel solved a number of mysterious problems, such as how the ribosome ensured proper insertion of membrane proteins into a lipid bilayer by direct docking of the tunnel exit with the translocon, as well as understanding the context in which proteins emerge from the ribosome and begin to fold.

Beyond Overall Morphology

Despite the characterization of essential features, going beyond the overall morphology of the ribosome proved to be difficult. Various approaches to chipping away at the problem resulted in a mass of data on protein-RNA interactions (Stern et al., 1989; Powers and Noller, 1995; Greuer et al., 1987; Gulle et al., 1988) and on the approximate spatial location of proteins by neutron scattering (Capel et al., 1987). The hope was that in conjunction with emerging structures of individual ribosomal proteins (Ramakrishnan and White, 1998) and improving cryoEM reconstructions, this information could be combined to generate a molecular model for the ribosome. There appeared to be no realistic alternative on the horizon. However, the low-resolution of the information and lack of an overall three-dimensional context made it difficult to obtain a model of sufficient accuracy and resolution to make deductions about its mechanism.

Crystallography of the Ribosome

Ever since the 1950s, crystallography has been used to determine the structure of macromolecules. Beginning in the 1960s (Byers, 1966), the ribosome itself had been shown to form two-dimensional crystalline sheets, and such sheets were used to obtain structures by electron microscopy (Unwin, 1977). These early studies showed that the ribosome had a defined structure that might be capable of producing three-dimensional crystals. Moreover, such crystals had been obtained for large viruses, the nucleosome core particle and F1 ATPase. Against this backdrop, it seemed reasonable to attempt the crystallization of the ribosome. Although the structure of some viruses had been determined to high resolution, they represented a special case of high symmetry. In contrast, even if suitable crystals could be obtained, solving the structure of such a large asymmetric structure as the ribosome seemed a pipe dream because the technology and methods to solve it did not exist.

Nevertheless, crystals of the 50S subunit were obtained from a thermophile, *Bacillus stearothermophilus* (Yonath et al., 1980), and were followed a few years later by crystals of the 30S subunit and the entire 70S ribosome from *Thermus thermophilus* (Trakhanov et al., 1987). These early crystals did not diffract well. Almost a decade later, the discovery of crystals of the 50S subunit from the archaeon *Haloarcula marismortui* that could diffract to 3 Å resolution (von Böhlen et al., 1991) meant that at least in principle it was possible to determine an atomic resolution structure. Cryocrystallography, in which crystals are cooled to 100 K to minimize radiation damage, was another essential step that allowed data collection (Hope, 1988). Nevertheless, for many years, the prospect of solving the ribosome's structure or even producing maps with recognizable features seemed out of reach.

Ultimately, it took technical advances in synchrotron X-ray sources and detectors, computing, as well as the participation of several research groups who brought in new ideas for

structure determination, to eventually obtain complete atomic structures of the 50S and 30S subunits (Ban et al., 2000; Wimberly et al., 2000). These structures were used to model a 5.5 Å structure of the entire ribosome with mRNA and tRNAs the following year (Yusupov et al., 2001). A few years later, high-resolution structures were obtained of both the empty ribosome (Schuwirth et al., 2005) and that of a complex with mRNA and tRNAs (Selmer et al., 2006). Currently there are crystal structures of the bacterial ribosome in many functional states (reviewed in Voorhees and Ramakrishnan, 2013). Moreover, there are now high-resolution structures of the entire eukaryotic ribosome (Ben-Shem et al., 2011) and the isolated 40S and 60S subunits (Rabl et al., 2011; Klinge et al., 2011).

These structures have transformed our understanding of the ribosome. The peptidyl transferase and decoding centers consist almost entirely of rRNA, showing that the ribosome is fundamentally an RNA-based enzyme. The structures have enabled sophisticated biochemical and genetic experiments to elucidate both basic mechanisms and the regulation of ribosome function. Moreover, the structures have enabled the direct visualization of antibiotics bound to the ribosome, enabling a better understanding of their mode of action and the potential for design of new, improved antibiotics (reviewed in Wilson, 2014).

The Ribosome in the 21st Century

How Does the Ribosome Work?

Whereas biochemical experiments designed to obtain structural information, like footprinting and crosslinking, have been almost completely superseded by direct methods such as crystallography or cryoEM, those designed to probe mechanisms are thriving. In particular, the development of pre-steady-state kinetics that use a combination of fluorescent reporters and quench flow methods has greatly helped dissect the various steps along the translation pathway (e.g., Rodnina and Wintermeyer, 2001). These methods have been further extended by the development of single-molecule fluorescence techniques, which can probe the rate and sequence of specific conformational changes in the ribosome during translation (Blanchard et al., 2004). Complementing such studies are the direct measurement of force (and thus work) required during translation (Liu et al., 2014).

Complementing the biochemistry, molecular dynamics, which has the potential to calculate both reaction mechanisms (Aqvist et al., 2012) and rates as well as trajectories in large-scale changes (Sanbonmatsu et al., 2005), is likely to play an increasing role in complementing experimental methods to probe ribosome function. These diverse methods are providing a detailed understanding of the mechanisms of translation.

In addition to understanding the internal mechanism of ribosomes, we are now getting insight into broader questions of translation in the cell. A particularly exciting advance is the method of ribosome profiling, which provides a genome-wide in vivo snapshot at nucleotide resolution of ribosomes along mRNA (Ingolia et al., 2009). This approach has already led to many major findings about gene expression in the cell, including translational pausing and the effect of various regulators of translation.

The Ribosome Is a Dynamic Rather Than Static Structure

That the ribosome cannot be static was understood right from the outset because during translocation after peptidyl transfer, the tRNAs and mRNA must move by precisely one codon with respect to the ribosome in order to allow a new cycle of elongation. Bretscher had proposed long ago that tRNAs move in two steps (Bretscher, 1968), first with respect to one subunit to form “hybrid” states and then with respect to the other, possibly coupled by a movement of the two subunits relative to each other. This hypothesis also provided a rationale for the existence of two subunits of ribosomes in all species. However, it languished for almost two decades until it was shown decisively that tRNAs move first with respect to the large subunit and then to the small (Moazed and Noller, 1989b).

Translocation could involve not just a movement of tRNAs but also a coupled rotation of the two subunits. Such a rotation was indeed observed by cryoEM (Frank and Agrawal, 2000). Moreover, even steps that were previously thought to be localized, such as decoding, which involves selection of a cognate tRNA corresponding to a codon, were shown to involve an induced conformational change (Pape et al., 1999). It is now known that major conformational changes are associated with virtually every step of translation including decoding and peptidyl transfer. Moreover, beginning with a study on tRNA associated with EF-Tu (Valle et al., 2002), studies have shown that the tRNAs themselves not only move, but distort considerably as they go through the ribosome.

Recent Advances in CryoEM Are Revolutionizing Structural Studies on the Ribosome

The development of cryoEM for asymmetric structures made a major impact right from the first structures of a ribosomal functional complex (Frank et al., 1995). Initially, the resolutions were not sufficient to derive molecular details without prior information, but they were nevertheless valuable to obtain structures of many states of the ribosome that could not be crystallized and also to analyze conformational changes. Recently however, as a result of direct electron detectors and improved software, it is possible to obtain near-atomic structures of the ribosome from only ~30,000 particles (Bai et al., 2013). Moreover, a complete atomic structure of the yeast mitochondrial large subunit has been determined de novo by cryoEM even when there were no known structural homologs for about half of the constituent proteins (Amunts et al., 2014). CryoEM is revolutionizing the structural biology of large complexes. Apart from eliminating the need for crystals, the method requires orders of magnitude less material, and most importantly can computationally sort both biochemical and conformational heterogeneity in a sample. It is thus ideal for studying complexes that are unstable or dynamic. For translation, this method will help solve more complex problems such as the structures of eukaryotic initiation complexes, of ribosomal complexes with membrane-bound receptors as well as with factors involved in quality control. It is safe to predict that for large complexes, cryoEM will largely supersede crystallography.

Conclusions

Those of us who began studying translation several decades ago could not have dreamed of the state of our understanding today

or the dizzying array of technologies that allow us to tackle harder problems in greater detail. We are now in position not only to understand the mechanism of the ribosome itself, but also how it is made, assembled, and regulated. We can visualize its interaction with quality control factors in the cell, see how viruses hijack it, and understand whether ribosomes are specialized for various functions. It is possible that 40 years hence our understanding of translation and its regulation will be as amazing to us now as today's knowledge would have been 40 years ago.

ACKNOWLEDGMENTS

Work in my laboratory is supported by the UK Medical Research Council (MC_U105184332) and the Wellcome Trust (WT096570). I thank James Lake for Figure 1A and Rebecca Voorhees for Figure 1B.

REFERENCES

- Agrawal, R.K., Penczek, P., Grassucci, R.A., Li, Y., Leith, A., Nierhaus, K.H., and Frank, J. (1996). Direct visualization of A-, P-, and E-site transfer RNAs in the *Escherichia coli* ribosome. *Science* 271, 1000–1002.
- Amunts, A., Brown, A., Bai, X.C., Ll  cer, J.L., Hussain, T., Emsley, P., Long, F., Murshudov, G., Scheres, S.H., and Ramakrishnan, V. (2014). Structure of the yeast mitochondrial large ribosomal subunit. *Science* 343, 1485–1489.
- Aqvist, J., Lind, C., Sund, J., and Wallin, G. (2012). Bridging the gap between ribosome structure and biochemistry by mechanistic computations. *Curr. Opin. Struct. Biol.* 22, 815–823.
- Bai, X.C., Fernandez, I.S., McMullan, G., and Scheres, S.H. (2013). Ribosome structures to near-atomic resolution from thirty thousand cryo-EM particles. *Elife* 2, e00461.
- Ban, N., Nissen, P., Hansen, J., Moore, P.B., and Steitz, T.A. (2000). The complete atomic structure of the large ribosomal subunit at 2.4 Å resolution. *Science* 289, 905–920.
- Ben-Shem, A., Garreau de Loubresse, N., Melnikov, S., Jenner, L., Yusupova, G., and Yusupov, M. (2011). The structure of the eukaryotic ribosome at 3.0 Å resolution. *Science* 334, 1524–1529.
- Bernabeu, C., and Lake, J.A. (1982). Nascent polypeptide chains emerge from the exit domain of the large ribosomal subunit: immune mapping of the nascent chain. *Proc. Natl. Acad. Sci. USA* 79, 3111–3115.
- Blanchard, S.C., Gonzalez, R.L., Kim, H.D., Chu, S., and Puglisi, J.D. (2004). tRNA selection and kinetic proofreading in translation. *Nat. Struct. Mol. Biol.* 11, 1008–1014.
- Bretscher, M.S. (1968). Translocation in protein synthesis: a hybrid structure model. *Nature* 218, 675–677.
- Brosius, J., Palmer, M.L., Kennedy, P.J., and Noller, H.F. (1978). Complete nucleotide sequence of a 16S ribosomal RNA gene from *Escherichia coli*. *Proc. Natl. Acad. Sci. USA* 75, 4801–4805.
- Brosius, J., Dull, T.J., and Noller, H.F. (1980). Complete nucleotide sequence of a 23S ribosomal RNA gene from *Escherichia coli*. *Proc. Natl. Acad. Sci. USA* 77, 201–204.
- Byers, B. (1966). Ribosome crystallization induced in chick embryo tissues by hypothermia. *J. Cell Biol.* 30, C1–C6.
- Capel, M.S., Engelman, D.M., Freeborn, B.R., Kjeldgaard, M., Langer, J.A., Ramakrishnan, V., Schindler, D.G., Schneider, D.K., Schoenborn, B.P., Sillers, I.-Y., et al. (1987). A complete mapping of the proteins in the small ribosomal subunit of *Escherichia coli*. *Science* 238, 1403–1406.
- Crick, F.H.C. (1968). The origin of the genetic code. *J. Mol. Biol.* 38, 367–379.
- D  ring, T., Mitchell, P., Osswald, M., Bochkariov, D., and Brimacombe, R. (1994). The decoding region of 16S RNA; a cross-linking study of the ribosomal A, P and E sites using tRNA derivatized at position 32 in the anticodon loop. *EMBO J.* 13, 2677–2685.

- Frank, J., and Agrawal, R.K. (2000). A ratchet-like inter-subunit reorganization of the ribosome during translocation. *Nature* 406, 318–322.
- Frank, J., Zhu, J., Penczek, P., Li, Y., Srivastava, S., Verschoor, A., Radermacher, M., Grassucci, R., Lata, R.K., and Agrawal, R.K. (1995). A model of protein synthesis based on cryo-electron microscopy of the *E. coli* ribosome. *Nature* 376, 441–444.
- Glötz, C., and Brimacombe, R. (1980). An experimentally-derived model for the secondary structure of the 16S ribosomal RNA from *Escherichia coli*. *Nucleic Acids Res.* 8, 2377–2395.
- Greuer, B., Osswald, M., Brimacombe, R., and Stöffler, G. (1987). RNA-protein cross-linking in *Escherichia coli* 30S ribosomal subunits; determination of sites on 16S RNA that are cross-linked to proteins S3, S4, S7, S9, S10, S11, S17, S18 and S21 by treatment with bis-(2-chloroethyl)-methylamine. *Nucleic Acids Res.* 15, 3241–3255.
- Guerrier-Takada, C., Gardiner, K., Marsh, T., Pace, N., and Altman, S. (1983). The RNA moiety of ribonuclease P is the catalytic subunit of the enzyme. *Cell* 35, 849–857.
- Gulle, H., Hoppe, E., Osswald, M., Greuer, B., Brimacombe, R., and Stöffler, G. (1988). RNA-protein cross-linking in *Escherichia coli* 50S ribosomal subunits; determination of sites on 23S RNA that are cross-linked to proteins L2, L4, L24 and L27 by treatment with 2-iminothiolane. *Nucleic Acids Res.* 16, 815–832.
- Held, W.A., Mizushima, S., and Nomura, M. (1973). Reconstitution of *Escherichia coli* 30S ribosomal subunits from purified molecular components. *J. Biol. Chem.* 248, 5720–5730.
- Hope, H. (1988). Cryocrystallography of biological macromolecules: a generally applicable method. *Acta Crystallogr. B* 44, 22–26.
- Ingolia, N.T., Ghaemmaghami, S., Newman, J.R., and Weissman, J.S. (2009). Genome-wide analysis in vivo of translation with nucleotide resolution using ribosome profiling. *Science* 324, 218–223.
- Klinge, S., Voigts-Hoffmann, F., Leibundgut, M., Arpagaus, S., and Ban, N. (2011). Crystal structure of the eukaryotic 60S ribosomal subunit in complex with initiation factor 6. *Science* 334, 941–948.
- Lake, J.A. (1976). Ribosome structure determined by electron microscopy of *Escherichia coli* small subunits, large subunits and monomeric ribosomes. *J. Mol. Biol.* 105, 131–139.
- Liu, T., Kaplan, A., Alexander, L., Yan, S., Wen, J.D., Lancaster, L., Wickersham, C.E., Fredrick, K., Noller, H., Tinoco, I., and Bustamante, C.J. (2014). Direct measurement of the mechanical work during translocation by the ribosome. *Elife* 3, e03406.
- Milligan, R.A., and Unwin, P.N. (1986). Location of exit channel for nascent protein in 80S ribosome. *Nature* 319, 693–695.
- Moazed, D., and Noller, H.F. (1986). Transfer RNA shields specific nucleotides in 16S ribosomal RNA from attack by chemical probes. *Cell* 47, 985–994.
- Moazed, D., and Noller, H.F. (1987a). Chloramphenicol, erythromycin, carbomycin and vernamycin B protect overlapping sites in the peptidyl transferase region of 23S ribosomal RNA. *Biochimie* 69, 879–884.
- Moazed, D., and Noller, H.F. (1987b). Interaction of antibiotics with functional sites in 16S ribosomal RNA. *Nature* 327, 389–394.
- Moazed, D., and Noller, H.F. (1989a). Interaction of tRNA with 23S rRNA in the ribosomal A, P, and E sites. *Cell* 57, 585–597.
- Moazed, D., and Noller, H.F. (1989b). Intermediate states in the movement of transfer RNA in the ribosome. *Nature* 342, 142–148.
- Moore, P.B. (1988). The ribosome returns. *Nature* 331, 223–227.
- Nierhaus, K.H., and Dohme, F. (1974). Total reconstitution of functionally active 50S ribosomal subunits from *Escherichia coli*. *Proc. Natl. Acad. Sci. USA* 71, 4713–4717.
- Nissen, P., Hansen, J., Ban, N., Moore, P.B., and Steitz, T.A. (2000). The structural basis of ribosome activity in peptide bond synthesis. *Science* 289, 920–930.
- Noller, H.F., and Chaires, J.B. (1972). Functional modification of 16S ribosomal RNA by kethoxal. *Proc. Natl. Acad. Sci. USA* 69, 3115–3118.
- Noller, H.F., Hoffarth, V., and Zimniak, L. (1992). Unusual resistance of peptidyl transferase to protein extraction procedures. *Science* 256, 1416–1419.
- Nomura, M., Tissières, A., and Lengyel, P. (1974). *Ribosomes* (New York: Cold Spring Harbor Monograph Series).
- Pape, T., Wintermeyer, W., and Rodnina, M. (1999). Induced fit in initial selection and proofreading of aminoacyl-tRNA on the ribosome. *EMBO J.* 18, 3800–3807.
- Powers, T., and Noller, H.F. (1995). Hydroxyl radical footprinting of ribosomal proteins on 16S rRNA. *RNA* 1, 194–209.
- Rabl, J., Leibundgut, M., Ataide, S.F., Haag, A., and Ban, N. (2011). Crystal structure of the eukaryotic 40S ribosomal subunit in complex with initiation factor 1. *Science* 331, 730–736.
- Ramakrishnan, V., and White, S.W. (1998). Ribosomal protein structures: insights into the architecture, machinery and evolution of the ribosome. *Trends Biochem. Sci.* 23, 208–212.
- Rheinberger, H.J., Sternbach, H., and Nierhaus, K.H. (1981). Three tRNA binding sites on *Escherichia coli* ribosomes. *Proc. Natl. Acad. Sci. USA* 78, 5310–5314.
- Rodnina, M.V., and Wintermeyer, W. (2001). Fidelity of aminoacyl-tRNA selection on the ribosome: kinetic and structural mechanisms. *Annu. Rev. Biochem.* 70, 415–435.
- Sanbonmatsu, K.Y., Joseph, S., and Tung, C.S. (2005). Simulating movement of tRNA into the ribosome during decoding. *Proc. Natl. Acad. Sci. USA* 102, 15854–15859.
- Schuwirth, B.S., Borovinskaya, M.A., Hau, C.W., Zhang, W., Vila-Sanjurjo, A., Holton, J.M., and Cate, J.H. (2005). Structures of the bacterial ribosome at 3.5 Å resolution. *Science* 310, 827–834.
- Selmer, M., Dunham, C.M., Murphy, F.V.T., 4th, Weixlbaumer, A., Petry, S., Kelley, A.C., Weir, J.R., and Ramakrishnan, V. (2006). Structure of the 70S ribosome complexed with mRNA and tRNA. *Science* 313, 1935–1942.
- Shine, J., and Dalgarno, L. (1974). The 3'-terminal sequence of *Escherichia coli* 16S ribosomal RNA: complementarity to nonsense triplets and ribosome binding sites. *Proc. Natl. Acad. Sci. USA* 71, 1342–1346.
- Stark, H., Orlova, E.V., Rinke-Appel, J., Jünke, N., Mueller, F., Rodnina, M.V., Wintermeyer, W., Brimacombe, R., and van Heel, M. (1997). Arrangement of tRNAs in pre- and posttranslocational ribosomes revealed by electron cryomicroscopy. *Cell* 88, 19–28.
- Steitz, J.A. (1969). Polypeptide chain initiation: nucleotide sequences of the three ribosomal binding sites in bacteriophage R17 RNA. *Nature* 224, 957–964.
- Stern, S., Powers, T., Changchien, L.M., and Noller, H.F. (1989). RNA-protein interactions in 30S ribosomal subunits: folding and function of 16S rRNA. *Science* 244, 783–790.
- Trakhanov, S.D., Yusupov, M.M., Agalarov, S.C., Garber, M.B., Ryazantsev, S.N., Tischenko, S.V., and Shirokov, V.A. (1987). Crystallization of 70S ribosomes and 30S ribosomal subunits from *Thermus thermophilus*. *FEBS Lett.* 220, 319–322.
- Traub, P., and Nomura, M. (1968). Streptomycin resistance mutation in *Escherichia coli*: altered ribosomal protein. *Science* 160, 198–199.
- Unwin, P.N. (1977). Three-dimensional model of membrane-bound ribosomes obtained by electron microscopy. *Nature* 269, 118–122.
- Valle, M., Sengupta, J., Swami, N.K., Grassucci, R.A., Burkhardt, N., Nierhaus, K.H., Agrawal, R.K., and Frank, J. (2002). Cryo-EM reveals an active role for aminoacyl-tRNA in the accommodation process. *EMBO J.* 21, 3557–3567.
- von Böhlen, K., Makowski, I., Hansen, H.A.S., Bartels, H., Berkovitch-Yellin, Z., Zaytzev-Bashan, A., Meyer, S., Paulke, C., Franceschi, F., and Yonath, A. (1991). Characterization and preliminary attempts for derivatization of crystals of large ribosomal subunits from *Haloarcula marismortui* diffracting to 3 Å resolution. *J. Mol. Biol.* 222, 11–15.
- Voorhees, R.M., and Ramakrishnan, V. (2013). Structural basis of the translational elongation cycle. *Annu. Rev. Biochem.* 82, 203–236.

- Voorhees, R.M., Fernández, I.S., Scheres, S.H., and Hegde, R.S. (2014). Structure of the mammalian ribosome-Sec61 complex to 3.4 Å resolution. *Cell* 157, 1632–1643.
- Wettstein, F.O., and Noll, H. (1965). Binding of transfer ribonucleic acid to ribosomes engaged in protein synthesis: Number and properties of ribosomal binding sites. *J. Mol. Biol.* 11, 35–53.
- Wilson, D.N. (2014). Ribosome-targeting antibiotics and mechanisms of bacterial resistance. *Nat. Rev. Microbiol.* 12, 35–48.
- Wimberly, B.T., Brodersen, D.E., Clemons, W.M., Jr., Morgan-Warren, R.J., Carter, A.P., Vonrhein, C., Hartsch, T., and Ramakrishnan, V. (2000). Structure of the 30S ribosomal subunit. *Nature* 407, 327–339.
- Woese, C.R., Magrum, L.J., Gupta, R., Siegel, R.B., Stahl, D.A., Kop, J., Crawford, N., Brosius, J., Gutell, R., Hogan, J.J., and Noller, H.F. (1980). Secondary structure model for bacterial 16S ribosomal RNA: phylogenetic, enzymatic and chemical evidence. *Nucleic Acids Res.* 8, 2275–2293.
- Woese, C.R., Kandler, O., and Wheelis, M.L. (1990). Towards a natural system of organisms: proposal for the domains Archaea, Bacteria, and Eucarya. *Proc. Natl. Acad. Sci. USA* 87, 4576–4579.
- Yonath, A., Mussig, J., Tesche, B., Lorenz, S., Erdmann, V.A., and Wittmann, H.G. (1980). Crystallization of the large ribosomal subunits from *Bacillus stearothermophilus*. *Biochem. Int.* 1, 428–435.
- Yonath, A., Leonard, K.R., and Wittmann, H.G. (1987). A tunnel in the large ribosomal subunit revealed by three-dimensional image reconstruction. *Science* 236, 813–816.
- Yusupov, M.M., Yusupova, G.Z., Baucom, A., Lieberman, K., Earnest, T.N., Cate, J.H., and Noller, H.F. (2001). Crystal structure of the ribosome at 5.5 Å resolution. *Science* 292, 883–896.
- Zaug, A.J., and Cech, T.R. (1986). The intervening sequence RNA of Tetrahymena is an enzyme. *Science* 231, 470–475.

A Tale of Chromatin and Transcription in 100 Structures

Patrick Cramer^{1,*}

¹Department of Molecular Biology, Max Planck Institute for Biophysical Chemistry, Am Fassberg 11, 37077 Göttingen, Germany

*Correspondence: patrick.cramer@mpibpc.mpg.de

<http://dx.doi.org/10.1016/j.cell.2014.10.047>

To celebrate a century of X-ray crystallography, I describe how 100 crystal structures influenced chromatin and transcription research.

Introduction

When Max von Laue first illuminated a crystal with X-rays 100 years ago, it was unclear what the obtained diffraction pattern meant. William Henry Bragg and his son Lawrence, however, soon realized that X-ray diffraction provided information about the inner structure of crystals. After decades of elaboration, X-ray crystallography advanced to become the most widely used method for the determination of 3D structures. Its application to biological macromolecules fostered the development of molecular biology in the second half of the 20th century. Crystallography defined biological paradigms such as molecular recognition, enzymatic catalysis, and allosteric regulation.

Until about 30 years ago, researchers could still follow publication of all new crystal structures of biomolecules. But then the pace at which new structures were solved increased rapidly due to the advent of new enabling techniques. Proteins were obtained in recombinant form, and nucleic acids were synthesized in large quantities. Crystals were cryo-cooled to slow down radiation damage. Synchrotron X-ray sources improved, and fast X-ray detection devices emerged. Modern computers and better software for structure determination became available. By now, 100,000 structures have been deposited in the protein database. Many of these revealed the inner workings of molecular machines, allowing researchers to rationalize phenotypes of mutations and to engineer biological processes.

Crystal structures can be like landmarks. They can guide us on our way toward a better understanding of a biological process (Shi, 2014 [this issue of *Cell*]). Landmark structures not only disclose some of life's secrets, but they also open up new frontiers. Here, I describe many of the structures that I consider to be landmark structures in the biology of chromatin, transcription, and epigenetics. I hope the resulting list of about 100 crystallographic structures, along with several structures obtained by other methods, exemplifies how structural information influenced the community and led to new concepts.

How DNA Is Structured

The proposal of the double-helical structure of DNA relied on X-ray diffraction patterns of DNA fibers obtained in the middle of the last century (Watson and Crick, 1953) (Figure 1). The direct observation of a nucleic acid duplex, however, had to await the crystal structure of a transfer RNA (tRNA) from yeast in 1974 (Robertus et al., 1974) (Figure 1). The structure of a B-DNA

duplex was solved only after DNA synthesis methods became available (Wing et al., 1980) (Figure 1). Crystal structures of DNA in A-form (Shakke et al., 1981) and Z-form (Wang et al., 1979) highlighted the sequence-dependent conformational flexibility of DNA.

In eukaryotic nuclei, DNA is packaged with histone proteins into chromatin. The fundamental unit of chromatin, the nucleosome core particle, was elucidated structurally in 1984 at a resolution of 7 Å (Richmond et al., 1984). When the resolution reached 2.8 Å, a detailed view of the nucleosome emerged that revealed the DNA conformation and DNA interactions with histones (Luger et al., 1997). The nucleosome core structure confirmed the structure of the free histone octamer (Arents et al., 1991). It further showed that the histone protein tails emerged between and around DNA duplexes to become available for interactions with other nucleosomes or chromatin.

The structure of higher-order chromatin is dynamic, but a complex of four nucleosomes could be crystallized and showed two stacks of nucleosomes and DNA that zigzagged between them (Schalch et al., 2005). Electron microscopy revealed how such tetranucleosome units may be arranged within a 30 nm fiber of chromatin (Song et al., 2014) and provided evidence for helical order in such fibers (Scheffer et al., 2011). Another electron microscopic study provided an alternative fiber model (Robinson et al., 2006). These results explained how extended DNA molecules can be packaged in the cell nucleus but also provided models for how chromatin regulates the accessibility of genes and their transcription. Whereas a compact chromatin structure can cause gene repression, gene activation requires chromatin opening and assembly of the transcription machinery at the promoter.

How DNA Is Recognized

To enable transcription, cells use transcription factors that bind to specific DNA sites. The first crystal structures of transcription factors included the bacterial catabolite activator protein CAP (McKay and Steitz, 1981) (Figure 1) and the bacteriophage lambda proteins cro (Anderson et al., 1981) and repressor (Pabo and Lewis, 1982). These structures contained helix-turn-helix regions that were involved in DNA binding and led to the concept of DNA-binding protein motifs. The studies of the bacteriophage proteins required protein overexpression because these transcription factors could not be isolated from natural sources in quantities required for structural studies.

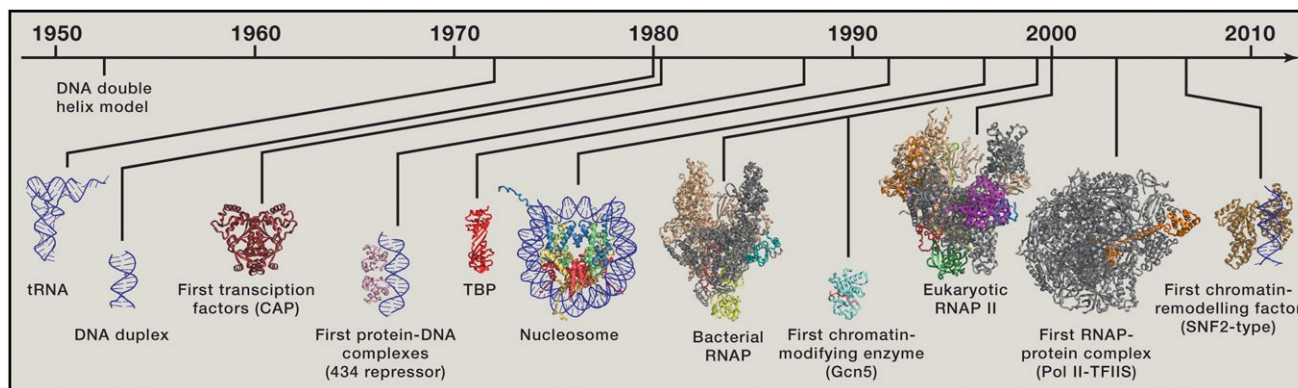


Figure 1. A Selection of Landmark Crystal Structures in the Biology of Chromatin and Transcription

From left to right, the depicted structures are yeast tRNA, a DNA duplex, the bacterial transcription factor CAP, the bacteriophage 434 repressor protein in complex with DNA, the eukaryotic TATA-binding protein TBP, the nucleosome, the bacterial RNA polymerase, the histone acetyltransferase Gcn5, the eukaryotic RNA polymerase 10-subunit core enzyme, the complete 12-subunit RNA polymerase II complex in complex with transcription factor TFIIS, and an archaeal Swi/Snf-type ATPase resembling the catalytic subunit found in many chromatin remodeling complexes. DNA is shown in blue, and proteins are depicted as ribbon models in different colors. For details, please refer to the text.

Structures of DNA-bound transcription factors led to the concept of sequence-specific DNA recognition. DNA complexes of the repressor proteins from bacteriophages 434 (Anderson et al., 1987) (Figure 1) and lambda (Jordan and Pabo, 1988) and of the 434 cro protein (Wolberger et al., 1988) revealed protein helices bound in the DNA major groove. DNA-binding helices were also observed in structures of homeodomains (Kissinger et al., 1990; Otting et al., 1990; Qian et al., 1989). In a “leucine zipper” of the GCN4 factor, protein helices in the DNA major groove were extended and used for factor dimerization (Ellenberger et al., 1992; König and Richmond, 1993). The transcription factors recognized target sequences via interactions of amino acid residues with DNA base edges. Such “direct readout” can be complemented by “indirect readout” of the DNA conformation via protein-DNA backbone contacts (Lesser et al., 1990; Otwinowski et al., 1988).

Later protein-DNA complex structures revealed a variety of DNA-binding structural motifs. The transcription factor NF- κ B uses a β barrel fold to contact DNA via protein loops (Becker et al., 1998; Ghosh et al., 1995; Müller et al., 1995). Transcription factors of the zinc finger family recognize DNA with small protein folds that are stabilized by zinc ions (Fairall et al., 1993; Luisi et al., 1991; Marmorstein et al., 1992; Pavletich and Pabo, 1991). Zinc fingers were later used for the design of proteins with new DNA-binding specificities (Choo et al., 1994). This catalyzed the development of protein and genome engineering as new research fields. Zinc fingers were also present in Klf4 (Schuetz et al., 2011), which, together with transcription factors Oct4, Sox2, and c-Myc, enables reprogramming of the genome and generation of induced pluripotent stem cells.

Crystallography also showed how transcription factors bind to adjacent DNA sites for combinatorial gene regulation. The DNA-bound structures of yeast MAT α 2 interacting with MAT α 1 (Li et al., 1995) and with MCM1 (Tan and Richmond, 1998) revealed factor-factor interactions that underlie synergistic DNA binding. This concept held for human transcription factors (Piper et al., 1999). Oct4 and Sox2 can also bind to neighboring DNA sites

(Reményi et al., 2003). Crystallography also led to a model of an “enhanceosome” containing eight transcription factors bound to DNA (Panne et al., 2007). Here, binding of one factor induces a DNA conformation that promotes binding of a neighboring factor.

Transcription factors can also bend DNA dramatically. The bacterial CAP protein bends DNA by 90 degrees to enable specific DNA recognition (Schultz et al., 1991). The eukaryotic TATA box-binding protein (TBP) also introduces a 90 degree bend into DNA (Kim et al., 1993a, 1993b). The integration host factor (Rice et al., 1996) and the mitochondrial transcription factor A (Ngo et al., 2011; Rubio-Cosials et al., 2011) can even bend DNA by 180 degrees, inducing a U-turn. To achieve DNA bending, proteins can use two strategies. They can insert amino acid residues like wedges between DNA base pairs to disrupt base stacking and may also neutralize backbone charges on one side of DNA, which induces bending due to the repulsion of phosphates on the opposite strand.

How DNA Binding Is Regulated

Crystallography further established concepts that explained bacterial gene regulation. In the *trp* operon, the Trp repressor protein inhibits expression of enzymes required for tryptophan biosynthesis when enough of the amino acid is available. The structure of the Trp repressor revealed a homodimer with a DNA-binding helix in each monomer (Schevitz et al., 1985). Binding of the regulator tryptophan alters the relative position of the two helices to enable DNA binding and gene repression (Otwinowski et al., 1988). These studies also showed that water molecules in the protein-DNA interface may contribute to sequence-specific DNA recognition.

Many transcription factors contain not only a DNA-binding domain but also additional domains that can activate transcription or bind other transcription factors or small-molecule regulators. In the bacterial *lac* operon, the Lac repressor protein binds DNA to control the expression of enzymes involved in lactose metabolism. The lac repressor contains a domain that binds

the regulator lactose (Friedman et al., 1995) and a domain that binds DNA (Bell and Lewis, 2000; Lewis et al., 1996). The regulator allosterically changes the orientation of the DNA-binding domains, leading to DNA dissociation and gene activation.

Eukaryotic transcription factors are generally also modular. The nuclear receptor for the hormone estrogen comprises a DNA-binding domain (Schwabe et al., 1993) and a hormone-binding domain (Brzozowski et al., 1997). Binding of estrogen influences receptor dimerization and its interaction with other factors that regulate transcription. The tumor suppressor p53 contains a tetramerization domain that clamps four p53 subunits together (Clore et al., 1994) and a DNA-binding domain that harbors tumorigenic mutations that can impair DNA binding (Cho et al., 1994).

Transcription factors are often regulated by nuclear localization. For example, the inhibitory protein I- κ B retains the transcription factor NF- κ B in the cytoplasm by masking its nuclear localization sequence (Huxford et al., 1998). External signals remove I- κ B, leading to nuclear import of NF- κ B, DNA binding, and gene activation. Similarly, external signals trigger phosphorylation of cytosolic STAT transcription factors (Becker et al., 1998), which leads to factor dimerization, nuclear import, DNA binding, and gene activation.

How DNA Directs RNA Synthesis: Single-Subunit RNA Polymerases

It took until the late 1990s to obtain a crystallographic view of transcription. The shape of the single-subunit DNA-dependent RNA polymerase from bacteriophage T7 was observed in a medium-resolution crystallographic study (Sousa et al., 1993). The crystal structure of T7 RNA polymerase revealed an active center cleft and a domain that binds promoter DNA and enables sequence-specific initiation of RNA synthesis (Jeruzalmi and Steitz, 1998; Cheetham et al., 1999). In the structure of a T7 RNA polymerase transcribing complex, the DNA template strand formed a hybrid duplex with a transcript of three nucleotides at the active site (Cheetham and Steitz, 1999). This suggested that the polymerase could only hold a three base pair DNA-RNA hybrid, but later structures of the T7 elongation complex revealed extensive refolding of the polymerase, which then accommodated a seven to eight base pair hybrid (Tahirov et al., 2002; Yin and Steitz, 2002).

These studies established many concepts in DNA-directed RNA synthesis. They highlighted the importance of promoter recognition and showed how ribonucleotide substrates are selected over deoxyribonucleotides to prevent synthesis of DNA. Addition of nucleotides to the RNA occurs in two steps by first binding the substrate nucleoside triphosphate to a preinsertion site and then moving it to the insertion site for catalysis (Temiakov et al., 2004). Details of RNA synthesis were revealed by X-ray studies of the single-subunit RNA polymerase from bacteriophage N4 (Basu and Murakami, 2013).

Eukaryotic cells also contain a single-subunit RNA polymerase, the polymerase transcribing the mitochondrial genome. Mitochondrial RNA polymerase structurally resembles T7 RNA polymerase but contains an additional region for promoter binding (Ringel et al., 2011). In contrast to T7 RNA polymerase, mitochondrial RNA polymerase, however, does not refold during the

transition from transcription initiation to elongation (Schwinghammer et al., 2013).

How DNA Is Transcribed in Cells: Multisubunit RNA Polymerases

The first structures of cellular RNA polymerases were obtained at the turn of the millennium. The structure of bacterial RNA polymerase from *Thermus aquaticus* (Zhang et al., 1999) was followed by the structure of yeast RNA polymerase II (Pol II), which synthesizes messenger RNA (Cramer et al., 2000, 2001) (Figure 1). The structures revealed functional elements and enabled structure-function analysis of cellular transcription. They also suggested that catalysis followed a two-metal ion mechanism (Cramer et al., 2000), as proposed for all nucleotide polymerases (Steitz and Steitz, 1993). Comparison of bacterial and eukaryotic RNA polymerases revealed a conserved multisubunit architecture and an active center cleft with a flexible bridge helix for the translocation of the polymerase relative to DNA.

The structure of a Pol II transcription elongation complex showed that the polymerase clamp domain closed over a DNA-RNA hybrid of eight to nine base pairs during transcription elongation and suggested the basis for nucleic acid strand separation during transcription (Gnatt et al., 2001). Later structures of bacterial RNA polymerase and Pol II transcription elongation complexes with bound nucleoside triphosphate substrate (Vasyliev et al., 2007; Wang et al., 2006) revealed that folding of the polymerase trigger loop closed the active site and suggested mechanisms of substrate selection. The same mechanisms likely occur in archaeal RNA polymerases, which show a remarkable similarity to Pol II (Hirata et al., 2008; Korkhin et al., 2009).

The first structure of Pol II with a bound transcription factor, the elongation factor TFIIS, showed that a single “tunable” active site was used both for RNA synthesis and RNA cleavage and indicated the mechanism of proofreading during transcription (Kettenberger et al., 2003) (Figure 1). Electron microscopy of an analogous bacterial complex revealed a similar topology (Opalka et al., 2003). The structure of an “arrested” Pol II elongation complex with a backtracked RNA provided insights into how TFIIS rescues polymerase that stalled during transcription (Cheung and Cramer, 2011). Crystal structures of a second eukaryotic RNA polymerase, Pol I, showed that a subunit corresponding to TFIIS was located at the active center of this enzyme (Engel et al., 2013; Fernández-Tornero et al., 2013).

Transcription is coordinated with cotranscriptional events such as RNA processing. This coordination is to a large extent achieved by binding of factors to the flexible C-terminal domain (CTD) of Pol II. Changes in CTD phosphorylation lead to an exchange of protein factors during transcription. The first structures of CTD peptides bound to CTD-binding proteins showed that the CTD adopts multiple conformations and revealed the basis for phosphorylation-specific binding (Fabrega et al., 2003; Meinhart and Cramer, 2004; Verdecia et al., 2000). Crystallography also revealed the mechanisms and the determinants for substrate specificity of CTD kinases (Baumli et al., 2012; Baumli et al., 2008; Böskén et al., 2014; Lolli et al., 2004; Schneider et al., 2011; Tahirov et al., 2010) and CTD phosphatases (Ghosh et al., 2008; Kamenski et al., 2004; Xiang et al., 2010).

How Transcription Starts

For the initiation of transcription, RNA polymerases cooperate with initiation factors to locate and open promoter DNA. The structure of the Pol II initiation factor TBP revealed a saddle-shaped molecule (Nikolov et al., 1992) (Figure 1) that bound the DNA minor groove and bent DNA by 90 degrees (Kim et al., 1993a, 1993b). The resulting TBP-DNA complex can recruit initiation factors TFIIA (Geiger et al., 1996; Tan et al., 1996) and TFIIB (Nikolov et al., 1995) on either side. Electron microscopy revealed the overall architectures of the large multiprotein initiation factors TFIID (Andel et al., 1999; Bieniossek et al., 2013; Cianfrocco et al., 2013; Leurent et al., 2002) and TFIIF (Chang and Kornberg, 2000; Gibbons et al., 2012; Schultz et al., 2000). Crystal structures for individual parts of TFIID, TFIIE, TFIIF, and TFIH were reported, including an archaeal homolog of the ATPase in TFIH (Fan et al., 2006) that functions in DNA opening.

An understanding of transcription initiation, however, had to await structures of initiation factor complexes with RNA polymerases. An initial structure of a partial Pol II-TFIIB complex (Bushnell et al., 2004) was consistent with crosslinking results (Chen and Hahn, 2003) and revealed a domain of TFIIB that bound Pol II to recruit it to the promoter. The structure of the Pol II-TFIIB complex (Kostrewa et al., 2009; Liu et al., 2010) enabled modeling of initiation complexes with closed double-stranded and open DNA. A subsequent structure of a Pol II-TFIIB complex with bound DNA and an initial RNA transcript showed that TFIIB alters the polymerase active site to allosterically activate RNA synthesis (Sainsbury et al., 2013).

Bacterial transcription initiation relies on sigma factors. The structure of a free sigma factor revealed its modular nature (Malhotra et al., 1996). Structures of RNA polymerase with bound sigma factor (Murakami et al., 2002b; Vassilyev et al., 2002) and with sigma factor and promoter DNA (Murakami et al., 2002a) showed that sigma factor bridges between polymerase and the promoter and suggested how sigma factor domains recognize DNA sequence elements. When promoter DNA is unwound, one sigma factor domain traps the nontemplate single strand of DNA (Feklistov and Darst, 2011). An alternative sigma factor is able to stabilize a flipped-out base from the nontemplate strand during DNA melting (Campagne et al., 2014). Structures of bacterial transcription initiation complexes showed how RNA polymerase and sigma factor cooperate to recognize promoter sequences, unwind DNA, and “preorganize” the template strand for RNA chain initiation (Zhang et al., 2012).

The initiation factors sigma and TFIIB perform similar functions, including DNA binding and opening and defining the start site of transcription. Comparison of bacterial and eukaryotic structures showed that sigma factor and TFIIB interact with roughly the same parts of their RNA polymerases but have unrelated folds, arguing for convergent evolution. Recently, electron microscopy and crosslinking provided the location of additional initiation factors in human and yeast Pol II initiation complexes (Grünberg et al., 2012; He et al., 2013; Mühlbacher et al., 2014; Murakami et al., 2013), enabling further crystallographic studies of transcription initiation.

How Transcription Initiation Is Regulated

Bacterial transcription can be regulated by direct interactions of transcription factors with the general transcription machinery and DNA. The CAP factor activates transcription by binding DNA and an adjacent domain of the polymerase, thereby recruiting the polymerase to the promoter (Benoff et al., 2002). Similarly, a protein of bacteriophage lambda activates transcription by binding to DNA and an adjacent domain of sigma (Jain et al., 2004). Eukaryotic transcription factors normally bind to coactivator complexes, which then bind the general transcription machinery. Coactivator binding requires activation domains, which are often unstructured in their free state but can adopt short helical structures upon coactivator binding (Brzovic et al., 2011; Kussie et al., 1996; Radhakrishnan et al., 1997). Vice versa, a coactivator can also form a α helix to bind a transcription factor (Shiau et al., 1998).

Many transcription factors bind to the coactivator complex Mediator, which consists of 25–35 subunits arranged in four modules. The crystal structure of the Mediator head module revealed an intricate fold with surfaces for interactions with Pol II and other Mediator modules (Imasaki et al., 2011; Larivière et al., 2012; Robinson et al., 2012). Recent electron microscopy revealed the central location of the head module within the overall Mediator architecture (Tsai et al., 2014; Wang et al., 2014). The mechanisms by which Mediator influences transcription remain to be explored, but there is evidence for a conformational change in Mediator induced by binding of a transcription activator (Meyer et al., 2010; Taatjes et al., 2002).

Transcription initiation is also regulated by methylation of DNA upstream of the promoter. In higher cells, this often occurs in CpG islands, which are DNA regions enriched for CpG dinucleotides. Hypermethylation of CpG islands generally leads to transcription repression. In a structure of a methyltransferase-DNA substrate complex, the target base was flipped out of the DNA double helix and inserted into the enzyme's active site (Klimasauskas et al., 1994). A human DNA methyltransferase uses additional domains to ensure that only hemimethylated CpG dinucleotides undergo methylation after replication, as seen in the enzyme-DNA complex structure (Song et al., 2011; Song et al., 2012).

How Chromatin Regulates Transcription

To make chromatin accessible for transcription, remodeling complexes use ATP hydrolysis to change nucleosome position and structure. Many remodelers contain a Swi/Snf family ATPase that induces DNA translocation with respect to histones (Dürr et al., 2005; Thomä et al., 2005) (Figure 1). A combination of structural techniques provided the architecture of chromatin remodelers ISW1a (Yamada et al., 2011), INO80 (Saravanan et al., 2012; Tosi et al., 2013), and SWR1 (Nguyen et al., 2013) and indicated how these machines bind nucleosomes, although the structure of the “remodeled nucleosome state” has remained elusive. Remodeling complexes not only bind nucleosomes—they can also be regulated by nucleosomes allosterically (Clapier and Cairns, 2012; Hauk et al., 2010; Racki et al., 2014).

Proper nucleosome assembly by histone chaperones is required to repress cryptic transcription that can produce aberrant RNAs from nonpromoter regions. X-ray studies unraveled

chaperone structures and showed that chaperones disrupt histone interfaces or mimic nucleosomal DNA to prevent promiscuous histone interactions prior to their assembly into nucleosomes (Elsässer et al., 2012; English et al., 2006; Hondele et al., 2013; Hu et al., 2011; Park and Luger, 2006). Certain assembly factors incorporate histone variants into nucleosomes at specific sites. The histone variant H2A.Z changes nucleosome structure at active promoters (Suto et al., 2000) and is removed by a specific chaperone (Obri et al., 2014), and nucleosomes containing the centromeric histone H3 variant are apparently destabilized compared to canonical nucleosomes (Tachiwana et al., 2011).

Structural studies also revealed details on how proteins recognize nucleosomes. A structure of a nucleosome in complex with RCC1, a regulator of chromatin condensation, showed how this protein recognizes both histones and DNA to specifically bind nucleosomes (Makde et al., 2010). Another nucleosome structure in complex with the gene-silencing factor Sir3 indicated how protein-nucleosome interactions are regulated by modifications in histone tails (Armache et al., 2011).

How Chromatin Marks Function

Covalent histone modifications provide another layer of gene regulation (Strahl and Allis, 2000). Histone modifications include acetylation, methylation, phosphorylation, and ubiquitination and can be associated with active transcription or repression. Enzymes that set or remove these marks are known as “writers” and “erasers,” respectively. The structure of the histone acetyltransferase Gcn5 provided insights into how chromatin marks are written (Rojas et al., 1999) (Figure 1). The structure of a portion of the acetyltransferase p300 explained mutations associated with human cancers (Liu et al., 2008). The subunit organization of the large acetyltransferase complex NuA4 and its mode of interaction with the nucleosome were revealed by electron microscopy (Chittuluru et al., 2011). Crystal structures of histone methyltransferases revealed their mechanism and how specificity for histones was achieved (Kwon et al., 2003; Min et al., 2002; Wilson et al., 2002; Xiao et al., 2003; Zhang et al., 2002).

Crystallography also showed how eraser enzymes work. The structure of a bacterial homolog of a histone deacetylase (Finnin et al., 1999) was followed by structures of the NAD-dependent Sir2 enzyme (Avalos et al., 2002; Finnin et al., 2001; Min et al., 2001) and eukaryotic zinc-dependent histone deacetylases (Somoza et al., 2004; Vannini et al., 2004). Histone demethylase structures of the LSD1 (Chen et al., 2006; Stavropoulos et al., 2006; Yang et al., 2006) and JmJ (Ng et al., 2007) classes revealed the basis for substrate specificity. The structure of the four-subunit deubiquitination module of the SAGA complex provided the basis for substrate specificity and activation of this eraser (Köhler et al., 2010; Samara et al., 2010).

Histone marks can recruit proteins via specific “reader” domains. The bromodomain binds acetylated lysine residues, as observed for the factors P/CAF, Taf1, and Gcn5 (Dhalluin et al., 1999; Jacobson et al., 2000; Owen et al., 2000). The chromodomain binds methylated lysines, as exemplified by the HP1 chromodomain bound to a histone H3 peptide methylated at lysine-9 (Jacobs and Khorasanizadeh, 2002). The PHD finger

domain also binds a methylated lysine residue by trapping it into an aromatic cage that is lined with residues mutated in cancer (Li et al., 2006; Peña et al., 2006). The pockets in reader and eraser proteins were consequently explored for drug design (Filippakopoulos et al., 2010). Multiple histone marks can be read by a single reader domain (Morinière et al., 2009) or by combinatorial binding of readers (Jacobson et al., 2000; Tsai et al., 2010; Xi et al., 2011). Multiple histone marks influence transcription activity, for example, via the initiation factor TFIID that binds histone tails with marks for active transcription (Vermeulen et al., 2007).

Toward Structural Cell Biology

In the coming years, X-ray crystallography will likely continue to provide landmark structures that elucidate unknown mechanisms in chromatin and transcription biology. However, many proteins that function in chromatin and transcription are modular by design. To resolve the structure of flexible factors and transient multicomponent complexes, crystallography will often be integrated with complementary techniques. Of particular importance will be electron microscopy, which enables placement of crystal structures of complex components but can now also reach high resolution that enables building of atomic models (Kühlbrandt, 2014; Wong et al., 2014). Crosslinking and mass spectrometry (Gingras et al., 2007; Serpa et al., 2012) will be routinely used to derive the relative position of known structures and to support models obtained by a combination of electron microscopy and crystallography.

A central future challenge will be the analysis of molecular structures within their cellular context and of structural changes in space and time. Advanced light microscopy techniques can now resolve detailed structures of assemblies such as nuclear pore complexes (Szymborska et al., 2013) or the cytoskeleton (Xu et al., 2013). A combination of *in vivo* crosslinking with deep sequencing and computer simulations can elucidate the overall folding of chromosomes (Naumova et al., 2013). Electron tomography provides three-dimensional images of the nuclear pore complex (Beck et al., 2004; Bui et al., 2013) or polysomes (Brandt et al., 2009).

We are witnessing the advent of a new research field that may be referred to as structural cell biology. Structural biologists may soon tackle most fundamental questions in biology. What is the conformational space for genomes, and how is it explored and utilized during gene activation and cell differentiation? What is the three-dimensional structure of genes and how does it change during transcription? What is the dynamic architecture of transient RNA assemblies with multiple proteins? Answers to these questions will require the development of new techniques and algorithms to bridge resolution gaps, to integrate structural data from multiple sources, and to embed structures into their biological context.

ACKNOWLEDGMENTS

I thank members of our laboratory and the following colleagues who provided comments: David Allis, Karim-Jean Armache, Alan Cheung, Richard Ebright, Andreas Ladurner, Robert Landick, Karolin Luger, Ronen Marmorstein, Christoph Müller, Dinshaw Patel, Alessandro Vannini, and Cynthia Wolberger. I thank Lucas Farnung for help with preparing the figure.

REFERENCES

- Andel, F., 3rd, Ladurner, A.G., Inouye, C., Tjian, R., and Nogales, E. (1999). Three-dimensional structure of the human TFIID-IIA-IIB complex. *Science* 286, 2153–2156.
- Anderson, W.F., Ohlendorf, D.H., Takeda, Y., and Matthews, B.W. (1981). Structure of the cro repressor from bacteriophage lambda and its interaction with DNA. *Nature* 290, 754–758.
- Anderson, J.E., Ptashne, M., and Harrison, S.C. (1987). Structure of the repressor-operator complex of bacteriophage 434. *Nature* 326, 846–852.
- Arents, G., Burlingame, R.W., Wang, B.C., Love, W.E., and Moudrianakis, E.N. (1991). The nucleosomal core histone octamer at 3.1 Å resolution: a tripartite protein assembly and a left-handed superhelix. *Proc. Natl. Acad. Sci. USA* 88, 10148–10152.
- Armache, K.J., Garlick, J.D., Canzio, D., Narlikar, G.J., and Kingston, R.E. (2011). Structural basis of silencing: Sir3 BAH domain in complex with a nucleosome at 3.0 Å resolution. *Science* 334, 977–982.
- Avalos, J.L., Celic, I., Muhammad, S., Cosgrove, M.S., Boeke, J.D., and Wolberger, C. (2002). Structure of a Sir2 enzyme bound to an acetylated p53 peptide. *Mol. Cell* 10, 523–535.
- Basu, R.S., and Murakami, K.S. (2013). Watching the bacteriophage N4 RNA polymerase transcription by time-dependent soak-trigger-freeze X-ray crystallography. *J. Biol. Chem.* 288, 3305–3311.
- Baumli, S., Lolli, G., Lowe, E.D., Troiani, S., Rusconi, L., Bullock, A.N., Debreczeni, J.E., Knapp, S., and Johnson, L.N. (2008). The structure of P-TEFb (CDK9/cyclin T1), its complex with flavopiridol and regulation by phosphorylation. *EMBO J.* 27, 1907–1918.
- Baumli, S., Hole, A.J., Wang, L.Z., Noble, M.E., and Endicott, J.A. (2012). The CDK9 tail determines the reaction pathway of positive transcription elongation factor b. *Structure* 20, 1788–1795.
- Beck, M., Förster, F., Ecke, M., Plitzko, J.M., Melchior, F., Gerisch, G., Baummeister, W., and Medalia, O. (2004). Nuclear pore complex structure and dynamics revealed by cryoelectron tomography. *Science* 306, 1387–1390.
- Becker, S., Groner, B., and Müller, C.W. (1998). Three-dimensional structure of the Stat3beta homodimer bound to DNA. *Nature* 394, 145–151.
- Bell, C.E., and Lewis, M. (2000). A closer view of the conformation of the Lac repressor bound to operator. *Nat. Struct. Biol.* 7, 209–214.
- Benoff, B., Yang, H., Lawson, C.L., Parkinson, G., Liu, J., Blatter, E., Ebright, Y.W., Berman, H.M., and Ebright, R.H. (2002). Structural basis of transcription activation: the CAP-alpha CTD-DNA complex. *Science* 297, 1562–1566.
- Bieniossek, C., Papai, G., Schaffitzel, C., Garzoni, F., Chaillet, M., Scheer, E., Papadopoulos, P., Tora, L., Schultz, P., and Berger, I. (2013). The architecture of human general transcription factor TFIID core complex. *Nature* 493, 699–702.
- Bösken, C.A., Farnung, L., Hintermair, C., Merzel Schachter, M., Vogel-Bachmayr, K., Blazek, D., Anand, K., Fisher, R.P., Eick, D., and Geyer, M. (2014). The structure and substrate specificity of human Cdk12/Cyclin K. *Nat. Commun.* 5, 3505.
- Brandt, F., Etchells, S.A., Ortiz, J.O., Elcock, A.H., Hartl, F.U., and Baumeister, W. (2009). The native 3D organization of bacterial polysomes. *Cell* 136, 261–271.
- Brzovic, P.S., Heikaus, C.C., Kisselev, L., Vernon, R., Herbig, E., Pacheco, D., Warfield, L., Littlefield, P., Baker, D., Kleivit, R.E., and Hahn, S. (2011). The acidic transcription activator Gcn4 binds the mediator subunit Gal11/Med15 using a simple protein interface forming a fuzzy complex. *Mol. Cell* 44, 942–953.
- Brzozowski, A.M., Pike, A.C., Dauter, Z., Hubbard, R.E., Bonn, T., Engström, O., Ohman, L., Greene, G.L., Gustafsson, J.A., and Carlquist, M. (1997). Molecular basis of agonism and antagonism in the oestrogen receptor. *Nature* 389, 753–758.
- Bui, K.H., von Appen, A., DiGiulio, A.L., Ori, A., Sparks, L., Mackmull, M.T., Bock, T., Hagen, W., Andrés-Pons, A., Glavy, J.S., and Beck, M. (2013). Integrated structural analysis of the human nuclear pore complex scaffold. *Cell* 155, 1233–1243.
- Bushnell, D.A., Westover, K.D., Davis, R.E., and Kornberg, R.D. (2004). Structural basis of transcription: an RNA polymerase II-TFIIB cocrystal at 4.5 Å. *Science* 303, 983–988.
- Campagne, S., Marsh, M.E., Capitani, G., Vorholt, J.A., and Allain, F.H. (2014). Structural basis for -10 promoter element melting by environmentally induced sigma factors. *Nat. Struct. Mol. Biol.* 21, 269–276.
- Chang, W.H., and Kornberg, R.D. (2000). Electron crystal structure of the transcription factor and DNA repair complex, core TFIIF. *Cell* 102, 609–613.
- Cheetham, G.M., and Steitz, T.A. (1999). Structure of a transcribing T7 RNA polymerase initiation complex. *Science* 286, 2305–2309.
- Cheetham, G.M., Jeruzalmi, D., and Steitz, T.A. (1999). Structural basis for initiation of transcription from an RNA polymerase-promoter complex. *Nature* 399, 80–83.
- Chen, H.T., and Hahn, S. (2003). Binding of TFIIB to RNA polymerase II: Mapping the binding site for the TFIIB zinc ribbon domain within the preinitiation complex. *Mol. Cell* 12, 437–447.
- Chen, Y., Yang, Y., Wang, F., Wan, K., Yamane, K., Zhang, Y., and Lei, M. (2006). Crystal structure of human histone lysine-specific demethylase 1 (LSD1). *Proc. Natl. Acad. Sci. USA* 103, 13956–13961.
- Cheung, A.C., and Cramer, P. (2011). Structural basis of RNA polymerase II backtracking, arrest and reactivation. *Nature* 471, 249–253.
- Chittuluru, J.R., Chaban, Y., Monnet-Saksouk, J., Carrozza, M.J., Sapountzi, V., Selleck, W., Huang, J., Utley, R.T., Cramet, M., Allard, S., et al. (2011). Structure and nucleosome interaction of the yeast NuA4 and Piccolo-NuA4 histone acetyltransferase complexes. *Nat. Struct. Mol. Biol.* 18, 1196–1203.
- Cho, Y., Gorina, S., Jeffrey, P.D., and Pavletich, N.P. (1994). Crystal structure of a p53 tumor suppressor-DNA complex: understanding tumorigenic mutations. *Science* 265, 346–355.
- Choo, Y., Sánchez-García, I., and Klug, A. (1994). In vivo repression by a site-specific DNA-binding protein designed against an oncogenic sequence. *Nature* 372, 642–645.
- Cianfrocco, M.A., Kassavetis, G.A., Grob, P., Fang, J., Juven-Gershon, T., Kadonaga, J.T., and Nogales, E. (2013). Human TFIID binds to core promoter DNA in a reorganized structural state. *Cell* 152, 120–131.
- Clapier, C.R., and Cairns, B.R. (2012). Regulation of ISWI involves inhibitory modules antagonized by nucleosomal epitopes. *Nature* 492, 280–284.
- Clore, G.M., Omichinski, J.G., Sakaguchi, K., Zambrano, N., Sakamoto, H., Appella, E., and Gronenborn, A.M. (1994). High-resolution structure of the oligomerization domain of p53 by multidimensional NMR. *Science* 265, 386–391.
- Cramer, P., Bushnell, D.A., Fu, J., Gnat, A.L., Maier-Davis, B., Thompson, N.E., Burgess, R.R., Edwards, A.M., David, P.R., and Kornberg, R.D. (2000). Architecture of RNA polymerase II and implications for the transcription mechanism. *Science* 288, 640–649.
- Cramer, P., Bushnell, D.A., and Kornberg, R.D. (2001). Structural basis of transcription: RNA polymerase II at 2.8 Å resolution. *Science* 292, 1863–1876.
- Dhalluin, C., Carlson, J.E., Zeng, L., He, C., Aggarwal, A.K., and Zhou, M.M. (1999). Structure and ligand of a histone acetyltransferase bromodomain. *Nature* 399, 491–496.
- Dürr, H., Körner, C., Müller, M., Hickmann, V., and Hopfner, K.P. (2005). X-ray structures of the Sulfolobus solfataricus SWI2/SNF2 ATPase core and its complex with DNA. *Cell* 121, 363–373.
- Ellenberger, T.E., Brandl, C.J., Struhl, K., and Harrison, S.C. (1992). The GCN4 basic region leucine zipper binds DNA as a dimer of uninterrupted alpha helices: crystal structure of the protein-DNA complex. *Cell* 71, 1223–1237.
- Eisässer, S.J., Huang, H., Lewis, P.W., Chin, J.W., Allis, C.D., and Patel, D.J. (2012). DAXX envelops a histone H3.3-H4 dimer for H3.3-specific recognition. *Nature* 491, 560–565.
- Engel, C., Sainsbury, S., Cheung, A.C., Kostrewa, D., and Cramer, P. (2013). RNA polymerase I structure and transcription regulation. *Nature* 502, 650–655.

- English, C.M., Adkins, M.W., Carson, J.J., Churchill, M.E., and Tyler, J.K. (2006). Structural basis for the histone chaperone activity of Asf1. *Cell* 127, 495–508.
- Fabrega, C., Shen, V., Shuman, S., and Lima, C.D. (2003). Structure of an mRNA capping enzyme bound to the phosphorylated carboxy-terminal domain of RNA polymerase II. *Mol. Cell* 11, 1549–1561.
- Fairall, L., Schwabe, J.W., Chapman, L., Finch, J.T., and Rhodes, D. (1993). The crystal structure of a two zinc-finger peptide reveals an extension to the rules for zinc-finger/DNA recognition. *Nature* 366, 483–487.
- Fan, L., Arvai, A.S., Cooper, P.K., Iwai, S., Hanaoka, F., and Tainer, J.A. (2006). Conserved XPB core structure and motifs for DNA unwinding: implications for pathway selection of transcription or excision repair. *Mol. Cell* 22, 27–37.
- Feklistov, A., and Darst, S.A. (2011). Structural basis for promoter-10 element recognition by the bacterial RNA polymerase σ subunit. *Cell* 147, 1257–1269.
- Fernández-Tornero, C., Moreno-Morcillo, M., Rashid, U.J., Taylor, N.M., Ruiz, F.M., Gruene, T., Legrand, P., Steuerwald, U., and Müller, C.W. (2013). Crystal structure of the 14-subunit RNA polymerase I. *Nature* 502, 644–649.
- Filippakopoulos, P., Qi, J., Picaud, S., Shen, Y., Smith, W.B., Fedorov, O., Morse, E.M., Keates, T., Hickman, T.T., Felletar, I., et al. (2010). Selective inhibition of BET bromodomains. *Nature* 468, 1067–1073.
- Finnin, M.S., Donigan, J.R., Cohen, A., Richon, V.M., Rifkin, R.A., Marks, P.A., Breslow, R., and Pavletich, N.P. (1999). Structures of a histone deacetylase homologue bound to the TSA and SAHA inhibitors. *Nature* 401, 188–193.
- Finnin, M.S., Donigan, J.R., and Pavletich, N.P. (2001). Structure of the histone deacetylase SIRT2. *Nat. Struct. Biol.* 8, 621–625.
- Friedman, A.M., Fischmann, T.O., and Steitz, T.A. (1995). Crystal structure of lac repressor core tetramer and its implications for DNA looping. *Science* 268, 1721–1727.
- Geiger, J.H., Hahn, S., Lee, S., and Sigler, P.B. (1996). Crystal structure of the yeast TFIIB/TBP/DNA complex. *Science* 272, 830–836.
- Ghosh, G., van Duyne, G., Ghosh, S., and Sigler, P.B. (1995). Structure of NF-kappa B p50 homodimer bound to a kappa B site. *Nature* 373, 303–310.
- Ghosh, A., Shuman, S., and Lima, C.D. (2008). The structure of Fcp1, an essential RNA polymerase II CTD phosphatase. *Mol. Cell* 32, 478–490.
- Gibbons, B.J., Brignole, E.J., Azubel, M., Murakami, K., Voss, N.R., Bushnell, D.A., Asturias, F.J., and Kornberg, R.D. (2012). Subunit architecture of general transcription factor TFIIB. *Proc. Natl. Acad. Sci. USA* 109, 1949–1954.
- Gingras, A.C., Gstaiger, M., Raught, B., and Aebersold, R. (2007). Analysis of protein complexes using mass spectrometry. *Nat. Rev. Mol. Cell Biol.* 8, 645–654.
- Gnatt, A.L., Cramer, P., Fu, J., Bushnell, D.A., and Kornberg, R.D. (2001). Structural basis of transcription: an RNA polymerase II elongation complex at 3.3 Å resolution. *Science* 292, 1876–1882.
- Grünberg, S., Warfield, L., and Hahn, S. (2012). Architecture of the RNA polymerase II preinitiation complex and mechanism of ATP-dependent promoter opening. *Nat. Struct. Mol. Biol.* 19, 788–796.
- Hauk, G., McKnight, J.N., Nodelman, I.M., and Bowman, G.D. (2010). The chromodomains of the Chd1 chromatin remodeler regulate DNA access to the ATPase motor. *Mol. Cell* 39, 711–723.
- He, Y., Fang, J., Taatjes, D.J., and Nogales, E. (2013). Structural visualization of key steps in human transcription initiation. *Nature* 495, 481–486.
- Hirata, A., Klein, B.J., and Murakami, K.S. (2008). The X-ray crystal structure of RNA polymerase from Archaea. *Nature* 451, 851–854.
- Hondele, M., Stuwe, T., Hassler, M., Halbach, F., Bowman, A., Zhang, E.T., Nijmeijer, B., Kotthoff, C., Rybin, V., Amlacher, S., et al. (2013). Structural basis of histone H2A-H2B recognition by the essential chaperone FACT. *Nature* 499, 111–114.
- Hu, H., Liu, Y., Wang, M., Fang, J., Huang, H., Yang, N., Li, Y., Wang, J., Yao, X., Shi, Y., et al. (2011). Structure of a CENP-A-histone H4 heterodimer in complex with chaperone HJURP. *Genes Dev.* 25, 901–906.
- Huxford, T., Huang, D.B., Malek, S., and Ghosh, G. (1998). The crystal structure of the I κ B α /NF- κ B complex reveals mechanisms of NF- κ B inactivation. *Cell* 95, 759–770.
- Imasaki, T., Calero, G., Cai, G., Tsai, K.L., Yamada, K., Cardelli, F., Erdjument-Bromage, H., Tempst, P., Berger, I., Kornberg, G.L., et al. (2011). Architecture of the Mediator head module. *Nature* 475, 240–243.
- Jacobs, S.A., and Khorasanizadeh, S. (2002). Structure of HP1 chromodomain bound to a lysine 9-methylated histone H3 tail. *Science* 295, 2080–2083.
- Jacobson, R.H., Ladurner, A.G., King, D.S., and Tjian, R. (2000). Structure and function of a human TAFII250 double bromodomain module. *Science* 288, 1422–1425.
- Jain, D., Nickels, B.E., Sun, L., Hochschild, A., and Darst, S.A. (2004). Structure of a ternary transcription activation complex. *Mol. Cell* 13, 45–53.
- Jeruzalmi, D., and Steitz, T.A. (1998). Structure of T7 RNA polymerase complexed to the transcriptional inhibitor T7 lysozyme. *EMBO J.* 17, 4101–4113.
- Jordan, S.R., and Pabo, C.O. (1988). Structure of the lambda complex at 2.5 Å resolution: details of the repressor-operator interactions. *Science* 242, 893–899.
- Kamenski, T., Heilmeyer, S., Meinhardt, A., and Cramer, P. (2004). Structure and mechanism of RNA polymerase II CTD phosphatases. *Mol. Cell* 15, 399–407.
- Kettenberger, H., Armache, K.J., and Cramer, P. (2003). Architecture of the RNA polymerase II-TFIIS complex and implications for mRNA cleavage. *Cell* 114, 347–357.
- Kim, J.L., Nikolov, D.B., and Burley, S.K. (1993a). Co-crystal structure of TBP recognizing the minor groove of a TATA element. *Nature* 365, 520–527.
- Kim, Y., Geiger, J.H., Hahn, S., and Sigler, P.B. (1993b). Crystal structure of a yeast TBP/TATA-box complex. *Nature* 365, 512–520.
- Kissinger, C.R., Liu, B.S., Martin-Blanco, E., Kornberg, T.B., and Pabo, C.O. (1990). Crystal structure of an engrailed homeodomain-DNA complex at 2.8 Å resolution: a framework for understanding homeodomain-DNA interactions. *Cell* 63, 579–590.
- Klimasauskas, S., Kumar, S., Roberts, R.J., and Cheng, X. (1994). HhaI methyltransferase flips its target base out of the DNA helix. *Cell* 76, 357–369.
- Köhler, A., Zimmerman, E., Schneider, M., Hurt, E., and Zheng, N. (2010). Structural basis for assembly and activation of the heterotetrameric SAGA histone H2B deubiquitinase module. *Cell* 141, 606–617.
- König, P., and Richmond, T.J. (1993). The X-ray structure of the GCN4-bZIP bound to ATF/CREB site DNA shows the complex depends on DNA flexibility. *J. Mol. Biol.* 233, 139–154.
- Korkhin, Y., Unligil, U.M., Littlefield, O., Nelson, P.J., Stuart, D.I., Sigler, P.B., Bell, S.D., and Abrescia, N.G. (2009). Evolution of complex RNA polymerases: the complete archaeal RNA polymerase structure. *PLoS Biol.* 7, e1000102.
- Kostrewa, D., Zeller, M.E., Armache, K.J., Seizl, M., Leike, K., Thomm, M., and Cramer, P. (2009). RNA polymerase II-TFIIB structure and mechanism of transcription initiation. *Nature* 462, 323–330.
- Kühlbrandt, W. (2014). Biochemistry. The resolution revolution. *Science* 343, 1443–1444.
- Kussie, P.H., Gorina, S., Marechal, V., Elenbaas, B., Moreau, J., Levine, A.J., and Pavletich, N.P. (1996). Structure of the MDM2 oncoprotein bound to the p53 tumor suppressor transactivation domain. *Science* 274, 948–953.
- Kwon, T., Chang, J.H., Kwak, E., Lee, C.W., Joachimiak, A., Kim, Y.C., Lee, J., and Cho, Y. (2003). Mechanism of histone lysine methyl transfer revealed by the structure of SET7/9-AdoMet. *EMBO J.* 22, 292–303.
- Larivière, L., Plaschka, C., Seizl, M., Wenzek, L., Kurth, F., and Cramer, P. (2012). Structure of the Mediator head module. *Nature* 492, 448–451.
- Lesser, D.R., Kurpiewski, M.R., and Jen-Jacobson, L. (1990). The energetic basis of specificity in the Eco RI endonuclease–DNA interaction. *Science* 250, 776–786.
- Leurent, C., Sanders, S., Ruhlmann, C., Mallouh, V., Weil, P.A., Kirschner, D.B., Tora, L., and Schultz, P. (2002). Mapping histone fold TAFs within yeast TFIID. *EMBO J.* 21, 3424–3433.

- Lewis, M., Chang, G., Horton, N.C., Kercher, M.A., Pace, H.C., Schumacher, M.A., Brennan, R.G., and Lu, P. (1996). Crystal structure of the lactose operon repressor and its complexes with DNA and inducer. *Science* 271, 1247–1254.
- Li, T., Stark, M.R., Johnson, A.D., and Wolberger, C. (1995). Crystal structure of the MATA1/MAT alpha 2 homeodomain heterodimer bound to DNA. *Science* 270, 262–269.
- Li, H., Ilin, S., Wang, W., Duncan, E.M., Wysocka, J., Allis, C.D., and Patel, D.J. (2006). Molecular basis for site-specific read-out of histone H3K4me3 by the BPTF PHD finger of NURF. *Nature* 442, 91–95.
- Liu, X., Wang, L., Zhao, K., Thompson, P.R., Hwang, Y., Marmorstein, R., and Cole, P.A. (2008). The structural basis of protein acetylation by the p300/CBP transcriptional coactivator. *Nature* 451, 846–850.
- Liu, X., Bushnell, D.A., Wang, D., Calero, G., and Kornberg, R.D. (2010). Structure of an RNA polymerase II-TFIIB complex and the transcription initiation mechanism. *Science* 327, 206–209.
- Lolli, G., Lowe, E.D., Brown, N.R., and Johnson, L.N. (2004). The crystal structure of human CDK7 and its protein recognition properties. *Structure* 12, 2067–2079.
- Luger, K., Mäder, A.W., Richmond, R.K., Sargent, D.F., and Richmond, T.J. (1997). Crystal structure of the nucleosome core particle at 2.8 Å resolution. *Nature* 389, 251–260.
- Luisi, B.F., Xu, W.X., Otwinowski, Z., Freedman, L.P., Yamamoto, K.R., and Sigler, P.B. (1991). Crystallographic analysis of the interaction of the glucocorticoid receptor with DNA. *Nature* 352, 497–505.
- Makde, R.D., England, J.R., Yennawar, H.P., and Tan, S. (2010). Structure of RCC1 chromatin factor bound to the nucleosome core particle. *Nature* 467, 562–566.
- Malhotra, A., Severinova, E., and Darst, S.A. (1996). Crystal structure of a sigma 70 subunit fragment from *E. coli* RNA polymerase. *Cell* 87, 127–136.
- Marmorstein, R., Carey, M., Ptashne, M., and Harrison, S.C. (1992). DNA recognition by GAL4: structure of a protein-DNA complex. *Nature* 356, 408–414.
- McKay, D.B., and Steitz, T.A. (1981). Structure of catabolite gene activator protein at 2.9 Å resolution suggests binding to left-handed B-DNA. *Nature* 290, 744–749.
- Meinhart, A., and Cramer, P. (2004). Recognition of RNA polymerase II carboxy-terminal domain by 3'-RNA-processing factors. *Nature* 430, 223–226.
- Meyer, K.D., Lin, S.C., Bernecky, C., Gao, Y., and Taatjes, D.J. (2010). p53 activates transcription by directing structural shifts in Mediator. *Nat. Struct. Mol. Biol.* 17, 753–760.
- Min, J., Landry, J., Sternglanz, R., and Xu, R.M. (2001). Crystal structure of a SIR2 homolog-NAD complex. *Cell* 105, 269–279.
- Min, J., Zhang, X., Cheng, X., Grewal, S.I., and Xu, R.M. (2002). Structure of the SET domain histone lysine methyltransferase Ctr4. *Nat. Struct. Biol.* 9, 828–832.
- Morinière, J., Rousseaux, S., Steuerwald, U., Soler-López, M., Curtet, S., Vitte, A.L., Govin, J., Gaucher, J., Sadoul, K., Hart, D.J., et al. (2009). Cooperative binding of two acetylation marks on a histone tail by a single bromodomain. *Nature* 461, 664–668.
- Mühlbacher, W., Sainsbury, S., Hemann, M., Hantsche, M., Neyer, S., Herzog, F., and Cramer, P. (2014). Conserved architecture of the core RNA polymerase II initiation complex. *Nat. Commun.* 5, 4310.
- Müller, C.W., Rey, F.A., Sodeoka, M., Verdine, G.L., and Harrison, S.C. (1995). Structure of the NF-kappa B p50 homodimer bound to DNA. *Nature* 373, 311–317.
- Murakami, K.S., Masuda, S., Campbell, E.A., Muzzin, O., and Darst, S.A. (2002a). Structural basis of transcription initiation: an RNA polymerase holoenzyme-DNA complex. *Science* 296, 1285–1290.
- Murakami, K.S., Masuda, S., and Darst, S.A. (2002b). Structural basis of transcription initiation: RNA polymerase holoenzyme at 4 Å resolution. *Science* 296, 1280–1284.
- Murakami, K., Emlund, H., Kalisman, N., Bushnell, D.A., Adams, C.M., Azubel, M., Emlund, D., Levi-Kalishman, Y., Liu, X., Gibbons, B.J., et al. (2013). Architecture of an RNA polymerase II transcription pre-initiation complex. *Science* 342, 1238724.
- Naumova, N., Imakaev, M., Fudenberg, G., Zhan, Y., Lajoie, B.R., Mirny, L.A., and Dekker, J. (2013). Organization of the mitotic chromosome. *Science* 342, 948–953.
- Ng, S.S., Kavanagh, K.L., McDonough, M.A., Butler, D., Pilka, E.S., Lienard, B.M., Bray, J.E., Savitsky, P., Gileadi, O., von Delft, F., et al. (2007). Crystal structures of histone demethylase JMJD2A reveal basis for substrate specificity. *Nature* 448, 87–91.
- Ngo, H.B., Kaiser, J.T., and Chan, D.C. (2011). The mitochondrial transcription and packaging factor Tfam imposes a U-turn on mitochondrial DNA. *Nat. Struct. Mol. Biol.* 18, 1290–1296.
- Nguyen, V.Q., Ranjan, A., Stengel, F., Wei, D., Aebersold, R., Wu, C., and Leschziner, A.E. (2013). Molecular architecture of the ATP-dependent chromatin-remodeling complex SWR1. *Cell* 154, 1220–1231.
- Nikolov, D.B., Hu, S.H., Lin, J., Gasch, A., Hoffmann, A., Horikoshi, M., Chua, N.H., Roeder, R.G., and Burley, S.K. (1992). Crystal structure of TFIID TATA-box binding protein. *Nature* 360, 40–46.
- Nikolov, D.B., Chen, H., Halay, E.D., Usheva, A.A., Hisatake, K., Lee, D.K., Roeder, R.G., and Burley, S.K. (1995). Crystal structure of a TFIIB-TBP-TATA-element ternary complex. *Nature* 377, 119–128.
- Obri, A., Ouarrhni, K., Papin, C., Diebold, M.L., Padmanabhan, K., Marek, M., Stoll, I., Roy, L., Reilly, P.T., Mak, T.W., et al. (2014). ANP32E is a histone chaperone that removes H2A.Z from chromatin. *Nature* 505, 648–653.
- Opalka, N., Chlenov, M., Chacon, P., Rice, W.J., Wriggers, W., and Darst, S.A. (2003). Structure and function of the transcription elongation factor GreB bound to bacterial RNA polymerase. *Cell* 114, 335–345.
- Otting, G., Qian, Y.Q., Billeter, M., Müller, M., Affolter, M., Gehring, W.J., and Wüthrich, K. (1990). Protein-DNA contacts in the structure of a homeodomain-DNA complex determined by nuclear magnetic resonance spectroscopy in solution. *EMBO J.* 9, 3085–3092.
- Otwinowski, Z., Schevitz, R.W., Zhang, R.G., Lawson, C.L., Joachimiak, A., Marmorstein, R.Q., Luisi, B.F., and Sigler, P.B. (1988). Crystal structure of trp repressor/operator complex at atomic resolution. *Nature* 335, 321–329.
- Owen, D.J., Ornaghi, P., Yang, J.C., Lowe, N., Evans, P.R., Ballario, P., Neuhaus, D., Filetici, P., and Travers, A.A. (2000). The structural basis for the recognition of acetylated histone H4 by the bromodomain of histone acetyltransferase gcn5p. *EMBO J.* 19, 6141–6149.
- Pabo, C.O., and Lewis, M. (1982). The operator-binding domain of lambda repressor: structure and DNA recognition. *Nature* 298, 443–447.
- Panne, D., Maniatis, T., and Harrison, S.C. (2007). An atomic model of the interferon-beta enhanceosome. *Cell* 129, 1111–1123.
- Park, Y.J., and Luger, K. (2006). The structure of nucleosome assembly protein 1. *Proc. Natl. Acad. Sci. USA* 103, 1248–1253.
- Pavletich, N.P., and Pabo, C.O. (1991). Zinc finger-DNA recognition: crystal structure of a Zif268-DNA complex at 2.1 Å. *Science* 252, 809–817.
- Peña, P.V., Davrazou, F., Shi, X., Walter, K.L., Verkhusha, V.V., Gozani, O., Zhao, R., and Kutateladze, T.G. (2006). Molecular mechanism of histone H3K4me3 recognition by plant homeodomain of ING2. *Nature* 442, 100–103.
- Piper, D.E., Batchelor, A.H., Chang, C.P., Cleary, M.L., and Wolberger, C. (1999). Structure of a HoxB1-Pbx1 heterodimer bound to DNA: role of the hexapeptide and a fourth homeodomain helix in complex formation. *Cell* 96, 587–597.
- Qian, Y.Q., Billeter, M., Otting, G., Müller, M., Gehring, W.J., and Wüthrich, K. (1989). The structure of the Antennapedia homeodomain determined by NMR spectroscopy in solution: comparison with prokaryotic repressors. *Cell* 59, 573–580.
- Racki, L.R., Naber, N., Pate, E., Leonard, J.D., Cooke, R., and Narlikar, G.J. (2014). The histone H4 tail regulates the conformation of the ATP-binding pocket in the SNF2h chromatin remodeling enzyme. *J. Mol. Biol.* 426, 2034–2044.

- Radhakrishnan, I., Pérez-Alvarado, G.C., Parker, D., Dyson, H.J., Montminy, M.R., and Wright, P.E. (1997). Solution structure of the KIX domain of CBP bound to the transactivation domain of CREB: a model for activator:coactivator interactions. *Cell* 91, 741–752.
- Reményi, A., Lins, K., Nissen, L.J., Reinbold, R., Schöler, H.R., and Wilmanns, M. (2003). Crystal structure of a POU/HMG/DNA ternary complex suggests differential assembly of Oct4 and Sox2 on two enhancers. *Genes Dev.* 17, 2048–2059.
- Rice, P.A., Yang, S., Mizuuchi, K., and Nash, H.A. (1996). Crystal structure of an IHF-DNA complex: a protein-induced DNA U-turn. *Cell* 87, 1295–1306.
- Richmond, T.J., Finch, J.T., Rushton, B., Rhodes, D., and Klug, A. (1984). Structure of the nucleosome core particle at 7 Å resolution. *Nature* 311, 532–537.
- Ringel, R., Sologub, M., Morozov, Y.I., Litonin, D., Cramer, P., and Temiakov, D. (2011). Structure of human mitochondrial RNA polymerase. *Nature* 478, 269–273.
- Robertus, J.D., Ladner, J.E., Finch, J.T., Rhodes, D., Brown, R.S., Clark, B.F., and Klug, A. (1974). Structure of yeast phenylalanine tRNA at 3 Å resolution. *Nature* 250, 546–551.
- Robinson, P.J., Fairall, L., Huynh, V.A., and Rhodes, D. (2006). EM measurements define the dimensions of the “30-nm” chromatin fiber: evidence for a compact, interdigitated structure. *Proc. Natl. Acad. Sci. USA* 103, 6506–6511.
- Robinson, P.J., Bushnell, D.A., Trnka, M.J., Burlingame, A.L., and Kornberg, R.D. (2012). Structure of the mediator head module bound to the carboxy-terminal domain of RNA polymerase II. *Proc. Natl. Acad. Sci. USA* 109, 17931–17935.
- Rojas, J.R., Trievel, R.C., Zhou, J., Mo, Y., Li, X., Berger, S.L., Allis, C.D., and Marmorstein, R. (1999). Structure of Tetrahymena GCN5 bound to coenzyme A and a histone H3 peptide. *Nature* 401, 93–98.
- Rubio-Cosials, A., Sidow, J.F., Jiménez-Menéndez, N., Fernández-Millán, P., Montoya, J., Jacobs, H.T., Coll, M., Bernadó, P., and Solà, M. (2011). Human mitochondrial transcription factor A induces a U-turn structure in the light strand promoter. *Nat. Struct. Mol. Biol.* 18, 1281–1289.
- Sainsbury, S., Niesser, J., and Cramer, P. (2013). Structure and function of the initially transcribing RNA polymerase II-TFIIB complex. *Nature* 493, 437–440.
- Samara, N.L., Datta, A.B., Berndsen, C.E., Zhang, X., Yao, T., Cohen, R.E., and Wolberger, C. (2010). Structural insights into the assembly and function of the SAGA deubiquitinating module. *Science* 328, 1025–1029.
- Saravanan, M., Wuerges, J., Bose, D., McCormack, E.A., Cook, N.J., Zhang, X., and Wigley, D.B. (2012). Interactions between the nucleosome histone core and Arp8 in the INO80 chromatin remodeling complex. *Proc. Natl. Acad. Sci. USA* 109, 20883–20888.
- Schalch, T., Duda, S., Sargent, D.F., and Richmond, T.J. (2005). X-ray structure of a tetranucleosome and its implications for the chromatin fibre. *Nature* 436, 138–141.
- Scheffer, M.P., Eltsov, M., and Frangakis, A.S. (2011). Evidence for short-range helical order in the 30-nm chromatin fibers of erythrocyte nuclei. *Proc. Natl. Acad. Sci. USA* 108, 16992–16997.
- Schevitz, R.W., Otwinowski, Z., Joachimiak, A., Lawson, C.L., and Sigler, P.B. (1985). The three-dimensional structure of trp repressor. *Nature* 317, 782–786.
- Schneider, E.V., Böttcher, J., Blaesse, M., Neumann, L., Huber, R., and Maszkos, K. (2011). The structure of CDK8/CycC implicates specificity in the CDK/cyclin family and reveals interaction with a deep pocket binder. *J. Mol. Biol.* 412, 251–266.
- Schuetz, A., Nana, D., Rose, C., Zocher, G., Milanovic, M., Koenigsmann, J., Blasig, R., Heinemann, U., and Carstanjen, D. (2011). The structure of the Klf4 DNA-binding domain links to self-renewal and macrophage differentiation. *Cell. Mol. Life Sci.* 68, 3121–3131.
- Schultz, S.C., Shields, G.C., and Steitz, T.A. (1991). Crystal structure of a CAP-DNA complex: the DNA is bent by 90 degrees. *Science* 253, 1001–1007.
- Schultz, P., Fribourg, S., Poterszman, A., Mallouh, V., Moras, D., and Egly, J.M. (2000). Molecular structure of human TFIIB. *Cell* 102, 599–607.
- Schwabe, J.W., Chapman, L., Finch, J.T., and Rhodes, D. (1993). The crystal structure of the estrogen receptor DNA-binding domain bound to DNA: how receptors discriminate between their response elements. *Cell* 75, 567–578.
- Schwinghammer, K., Cheung, A.C., Morozov, Y.I., Agaronyan, K., Temiakov, D., and Cramer, P. (2013). Structure of human mitochondrial RNA polymerase elongation complex. *Nat. Struct. Mol. Biol.* 20, 1298–1303.
- Serpa, J.J., Parker, C.E., Petrotchenko, E.V., Han, J., Pan, J., and Borchers, C.H. (2012). Mass spectrometry-based structural proteomics. *Eur. J. Mass Spectrom.* (Chichester, Eng.) 18, 251–267.
- Shakked, Z., Rabinovich, D., Cruse, W.B., Egert, E., Kennard, O., Sala, G., Salisbury, S.A., and Viswamitra, M.A. (1981). Crystalline A-dna: the X-ray analysis of the fragment d(G-G-T-A-T-A-C-C). *Proc. R. Soc. Lond. B Biol. Sci.* 213, 479–487.
- Shi, Y. (2014). A glimpse of structural biology through X-ray crystallography. *Cell* 159, this issue, 995–1014.
- Shiau, A.K., Barstad, D., Loria, P.M., Cheng, L., Kushner, P.J., Agard, D.A., and Greene, G.L. (1998). The structural basis of estrogen receptor/coactivator recognition and the antagonism of this interaction by tamoxifen. *Cell* 95, 927–937.
- Somoza, J.R., Skene, R.J., Katz, B.A., Mol, C., Ho, J.D., Jennings, A.J., Luong, C., Arvai, A., Buggy, J.J., Chi, E., et al. (2004). Structural snapshots of human HDAC8 provide insights into the class I histone deacetylases. *Structure* 12, 1325–1334.
- Song, J., Rechkoblit, O., Bestor, T.H., and Patel, D.J. (2011). Structure of DNMT1-DNA complex reveals a role for autoinhibition in maintenance DNA methylation. *Science* 331, 1036–1040.
- Song, J., Teplova, M., Ishibe-Murakami, S., and Patel, D.J. (2012). Structure-based mechanistic insights into DNMT1-mediated maintenance DNA methylation. *Science* 335, 709–712.
- Song, F., Chen, P., Sun, D., Wang, M., Dong, L., Liang, D., Xu, R.M., Zhu, P., and Li, G. (2014). Cryo-EM study of the chromatin fiber reveals a double helix twisted by tetranucleosomal units. *Science* 344, 376–380.
- Sousa, R., Chung, Y.J., Rose, J.P., and Wang, B.C. (1993). Crystal structure of bacteriophage T7 RNA polymerase at 3.3 Å resolution. *Nature* 364, 593–599.
- Stavropoulos, P., Blobel, G., and Hoelz, A. (2006). Crystal structure and mechanism of human lysine-specific demethylase-1. *Nat. Struct. Mol. Biol.* 13, 626–632.
- Steitz, T.A., and Steitz, J.A. (1993). A general two-metal-ion mechanism for catalytic RNA. *Proc. Natl. Acad. Sci. USA* 90, 6498–6502.
- Strahl, B.D., and Allis, C.D. (2000). The language of covalent histone modifications. *Nature* 403, 41–45.
- Suto, R.K., Clarkson, M.J., Tremethick, D.J., and Luger, K. (2000). Crystal structure of a nucleosome core particle containing the variant histone H2A.Z. *Nat. Struct. Biol.* 7, 1121–1124.
- Szymborska, A., de Marco, A., Daigle, N., Cordes, V.C., Briggs, J.A., and Ellenberg, J. (2013). Nuclear pore scaffold structure analyzed by super-resolution microscopy and particle averaging. *Science* 341, 655–658.
- Taatjes, D.J., Näär, A.M., Andel, F., 3rd, Nogales, E., and Tjian, R. (2002). Structure, function, and activator-induced conformations of the CRSP coactivator. *Science* 295, 1058–1062.
- Tachiwana, H., Kagawa, W., Shiga, T., Osakabe, A., Miya, Y., Saito, K., Hayashi-Takanaka, Y., Oda, T., Sato, M., Park, S.Y., et al. (2011). Crystal structure of the human centromeric nucleosome containing CENP-A. *Nature* 476, 232–235.
- Tahirov, T.H., Temiakov, D., Anikin, M., Patlan, V., McAllister, W.T., Vassilyev, D.G., and Yokoyama, S. (2002). Structure of a T7 RNA polymerase elongation complex at 2.9 Å resolution. *Nature* 420, 43–50.
- Tahirov, T.H., Babayeva, N.D., Varzavand, K., Cooper, J.J., Sedore, S.C., and Price, D.H. (2010). Crystal structure of HIV-1 Tat complexed with human P-TEFb. *Nature* 465, 747–751.
- Tan, S., and Richmond, T.J. (1998). Crystal structure of the yeast MAT α 2/MCM1/DNA ternary complex. *Nature* 391, 660–666.

- Tan, S., Hunziker, Y., Sargent, D.F., and Richmond, T.J. (1996). Crystal structure of a yeast TFIIB/TBP/DNA complex. *Nature* **381**, 127–151.
- Temiaikov, D., Patlan, V., Anikin, M., McAllister, W.T., Yokoyama, S., and Vassilyev, D.G. (2004). Structural basis for substrate selection by t7 RNA polymerase. *Cell* **116**, 381–391.
- Thomä, N.H., Czyzewski, B.K., Alexeev, A.A., Mazin, A.V., Kowalczykowski, S.C., and Pavletich, N.P. (2005). Structure of the SWI2/SNF2 chromatin-re-modeling domain of eukaryotic Rad54. *Nat. Struct. Mol. Biol.* **12**, 350–356.
- Tosi, A., Haas, C., Herzog, F., Gilmozzi, A., Berninghausen, O., Ungewickell, C., Gerhold, C.B., Lakomek, K., Aebersold, R., Beckmann, R., and Hopfner, K.P. (2013). Structure and subunit topology of the INO80 chromatin remodeler and its nucleosome complex. *Cell* **154**, 1207–1219.
- Tsai, W.W., Wang, Z., Yiu, T.T., Akdemir, K.C., Xia, W., Winter, S., Tsai, C.Y., Shi, X., Schwarzer, D., Plunkett, W., et al. (2010). TRIM24 links a non-canonical histone signature to breast cancer. *Nature* **468**, 927–932.
- Tsai, K.L., Tomomori-Sato, C., Sato, S., Conaway, R.C., Conaway, J.W., and Asturias, F.J. (2014). Subunit architecture and functional modular rearrangements of the transcriptional mediator complex. *Cell* **157**, 1430–1444.
- Vannini, A., Volpari, C., Filocamo, G., Casavola, E.C., Brunetti, M., Renzoni, D., Chakravarty, P., Paolini, C., De Francesco, R., Gallinari, P., et al. (2004). Crystal structure of a eukaryotic zinc-dependent histone deacetylase, human HDAC8, complexed with a hydroxamic acid inhibitor. *Proc. Natl. Acad. Sci. USA* **101**, 15064–15069.
- Vassilyev, D.G., Sekine, S., Laptchenko, O., Lee, J., Vassilyeva, M.N., Borukhov, S., and Yokoyama, S. (2002). Crystal structure of a bacterial RNA polymerase holoenzyme at 2.6 Å resolution. *Nature* **417**, 712–719.
- Vassilyev, D.G., Vassilyeva, M.N., Zhang, J., Palangat, M., Artsimovitch, I., and Landick, R. (2007). Structural basis for substrate loading in bacterial RNA polymerase. *Nature* **448**, 163–168.
- Verdecia, M.A., Bowman, M.E., Lu, K.P., Hunter, T., and Noel, J.P. (2000). Structural basis for phosphoserine-proline recognition by group IV WW domains. *Nat. Struct. Biol.* **7**, 639–643.
- Vermeulen, M., Mulder, K.W., Denissov, S., Pijnappel, W.W., van Schaik, F.M., Varier, R.A., Baltissen, M.P., Stunnenberg, H.G., Mann, M., and Timmers, H.T. (2007). Selective anchoring of TFIID to nucleosomes by trimethylation of histone H3 lysine 4. *Cell* **131**, 58–69.
- Wang, A.H., Quigley, G.J., Kolpak, F.J., Crawford, J.L., van Boom, J.H., van der Marel, G., and Rich, A. (1979). Molecular structure of a left-handed double helical DNA fragment at atomic resolution. *Nature* **282**, 680–686.
- Wang, D., Bushnell, D.A., Westover, K.D., Kaplan, C.D., and Kornberg, R.D. (2006). Structural basis of transcription: role of the trigger loop in substrate specificity and catalysis. *Cell* **127**, 941–954.
- Wang, X., Sun, Q., Ding, Z., Ji, J., Wang, J., Kong, X., Yang, J., and Cai, G. (2014). Redefining the modular organization of the core Mediator complex. *Cell Res.* **24**, 796–808.
- Watson, J.D., and Crick, F.H. (1953). Molecular structure of nucleic acids; a structure for deoxyribose nucleic acid. *Nature* **171**, 737–738.
- Wilson, J.R., Jing, C., Walker, P.A., Martin, S.R., Howell, S.A., Blackburn, G.M., Gamblin, S.J., and Xiao, B. (2002). Crystal structure and functional analysis of the histone methyltransferase SET7/9. *Cell* **111**, 105–115.
- Wing, R., Drew, H., Takano, T., Broka, C., Tanaka, S., Itakura, K., and Dickerson, R.E. (1980). Crystal structure analysis of a complete turn of B-DNA. *Nature* **287**, 755–758.
- Wolberger, C., Dong, Y.C., Ptashne, M., and Harrison, S.C. (1988). Structure of a phage 434 Cro/DNA complex. *Nature* **335**, 789–795.
- Wong, W., Bai, X.C., Brown, A., Fernandez, I.S., Hanssen, E., Condrón, M., Tan, Y.H., Baum, J., and Scheres, S.H. (2014). Cryo-EM structure of the Plasmodium falciparum 80S ribosome bound to the anti-protozoan drug emetine. *eLife* **3**, e03080.
- Xi, Q., Wang, Z., Zaromytidou, A.I., Zhang, X.H., Chow-Tsang, L.F., Liu, J.X., Kim, H., Barlas, A., Manova-Todorova, K., Kaartinen, V., et al. (2011). A poised chromatin platform for TGF- β access to master regulators. *Cell* **147**, 1511–1524.
- Xiang, K., Nagaike, T., Xiang, S., Kilic, T., Beh, M.M., Manley, J.L., and Tong, L. (2010). Crystal structure of the human symplekin-Ssu72-CTD phosphopeptide complex. *Nature* **467**, 729–733.
- Xiao, B., Jing, C., Wilson, J.R., Walker, P.A., Vasisth, N., Kelly, G., Howell, S., Taylor, I.A., Blackburn, G.M., and Gamblin, S.J. (2003). Structure and catalytic mechanism of the human histone methyltransferase SET7/9. *Nature* **421**, 652–656.
- Xu, K., Zhong, G., and Zhuang, X. (2013). Actin, spectrin, and associated proteins form a periodic cytoskeletal structure in axons. *Science* **339**, 452–456.
- Yamada, K., Frouws, T.D., Angst, B., Fitzgerald, D.J., DeLuca, C., Schimmele, K., Sargent, D.F., and Richmond, T.J. (2011). Structure and mechanism of the chromatin remodelling factor ISW1a. *Nature* **472**, 448–453.
- Yang, M., Gocke, C.B., Luo, X., Borek, D., Tomchick, D.R., Machius, M., Otwinowski, Z., and Yu, H. (2006). Structural basis for CoREST-dependent demethylation of nucleosomes by the human LSD1 histone demethylase. *Mol. Cell* **23**, 377–387.
- Yin, Y.W., and Steitz, T.A. (2002). Structural basis for the transition from initiation to elongation transcription in T7 RNA polymerase. *Science* **298**, 1387–1395.
- Zhang, G., Campbell, E.A., Minakhin, L., Richter, C., Severinov, K., and Darst, S.A. (1999). Crystal structure of *Thermus aquaticus* core RNA polymerase at 3.3 Å resolution. *Cell* **98**, 811–824.
- Zhang, X., Tamaru, H., Khan, S.I., Horton, J.R., Keefe, L.J., Selker, E.U., and Cheng, X. (2002). Structure of the *Neurospora* SET domain protein DIM-5, a histone H3 lysine methyltransferase. *Cell* **111**, 117–127.
- Zhang, Y., Feng, Y., Chatterjee, S., Tuske, S., Ho, M.X., Arnold, E., and Ebright, R.H. (2012). Structural basis of transcription initiation. *Science* **338**, 1076–1080.

A Glimpse of Structural Biology through X-Ray Crystallography

Yigong Shi^{1,*}

¹Center for Structural Biology, Tsinghua-Peking Joint Center for Life Sciences, School of Life Sciences and School of Medicine, Tsinghua University, Beijing 100084, China

*Correspondence: shi-lab@tsinghua.edu.cn

<http://dx.doi.org/10.1016/j.cell.2014.10.051>

Since determination of the myoglobin structure in 1957, X-ray crystallography, as the anchoring tool of structural biology, has played an instrumental role in deciphering the secrets of life. Knowledge gained through X-ray crystallography has fundamentally advanced our views on cellular processes and greatly facilitated development of modern medicine. In this brief narrative, I describe my personal understanding of the evolution of structural biology through X-ray crystallography—using as examples mechanistic understanding of protein kinases and integral membrane proteins—and comment on the impact of technological development and outlook of X-ray crystallography.

Brief History

When Wilhelm Roentgen discovered X-ray in 1895, he could not have imagined the powerful applications of X-ray diffraction on crystals of biological samples. Max von Laue showed X-ray diffraction pattern of crystals in 1912, and William Lawrence Bragg derived a general equation, known as the Bragg's Law, to describe the founding principle of image formation by X-ray diffraction (Bragg, 1913) (Figure 1). James Sumner obtained the first crystal of jack bean urease in 1926 and showed the enzyme to be a protein (Figure 1). Max Perutz and John Kendrew decided to pursue crystal structures of proteins—hemoglobin and myoglobin—beginning in the 1940s at the Cavendish Laboratory, University of Cambridge. Their pioneering effort was encouraged by William Lawrence Bragg, who served as the Director of the Cavendish Laboratory between 1938 and 1954. In 1953, James Watson and Francis Crick, both employed at the Cavendish Laboratory, deduced a DNA double-helix model on the basis of X-ray fiber diffraction images of DNA generated by Rosalind Franklin (Watson and Crick, 1953).

The entire biological research community was both excited and shocked to see the very first crystal structure of a macromolecule in 1957—that of sperm whale myoglobin by John Kendrew (Kendrew et al., 1958). The structure of myoglobin, initially determined at 6 Å resolution but quickly improved to 2 Å (Kendrew et al., 1960), confirmed the α -helical conformation as proposed by Linus Pauling and Robert Corey (Pauling and Corey, 1951a, 1951b, 1951c; Pauling et al., 1951). Kendrew's success in structure determination of myoglobin was indispensably assisted by Perutz' solution to the phase problem—multiple isomorphous replacement through heavy atom soaks. Max Perutz presented his own X-ray structure on the larger protein hemoglobin at 5.5 Å (Perutz et al., 1960) and took a few years to improve the resolution to 2.8 Å (Perutz et al., 1968a, 1968b). Kendrew founded the *Journal of Molecular Biology* and served as Editor-in-Chief for a number of years. Kendrew also helped establish the European Molecular Biology Laboratory in Heidelberg and became its founding director. Perutz, on the other hand, founded and

directed the MRC Laboratory of Molecular Biology (Figure 1). Notably, the double-helix structure of DNA was finally visualized in 1980 by the X-ray structure of a 12-base-pair palindromic DNA, known as the Dickerson dodecamer (Wing et al., 1980).

DNA is the genetic material of almost all living matters, and proteins are the engines of life. Structural elucidation of DNA and protein is arguably the most important scientific discovery in the 20th century. Proposal of the double-helix structure of DNA has fundamentally changed our perception of life and has ushered in a new era of modern biology. Crystal structures of myoglobin and hemoglobin allowed us to link protein function to its chemical details. In many respects, the atomic details offered by X-ray crystallography allowed mechanistic understanding of protein function, which marks the beginning of molecular biology. Kendrew and Perutz have been fondly named fathers of molecular biology.

Early crystallographic studies focused on abundant proteins, most often enzymes, from animal organs and tissues. Following the successes on myoglobin and hemoglobin, structural information was obtained for at least seven additional proteins in the 1960s, including the first enzyme hen egg white lysozyme (Blake et al., 1965), ribonucleases A and S (Kartha et al., 1967; Wyckoff et al., 1967), chymotrypsin (Matthews et al., 1967), papain (Drenth et al., 1968), carboxypeptidase A (Lipscomb et al., 1969), and subtilisin (Wright et al., 1969). These structures, together with those of many other enzymes in the 1970s and beyond, reveal the active site conformations and catalytic mechanisms, which form the physical basis of molecular enzymology.

The Protein Data Bank (PDB), a central repository for three-dimensional structural data of macromolecules, was established in 1971 at the Brookhaven National Laboratory with seven initial entries. As of August 26, 2014, there were 102,863 total entries in the PDB, of which 88.7% were determined by X-ray crystallography, 10.3% by nuclear magnetic resonance (NMR), and 0.8% by electron microscopy (EM) (Figure 2A). Following structure determination of the lysozyme from bacteriophage T4 (T4 lysozyme) (Matthews and Remington, 1974), it became a paradigm

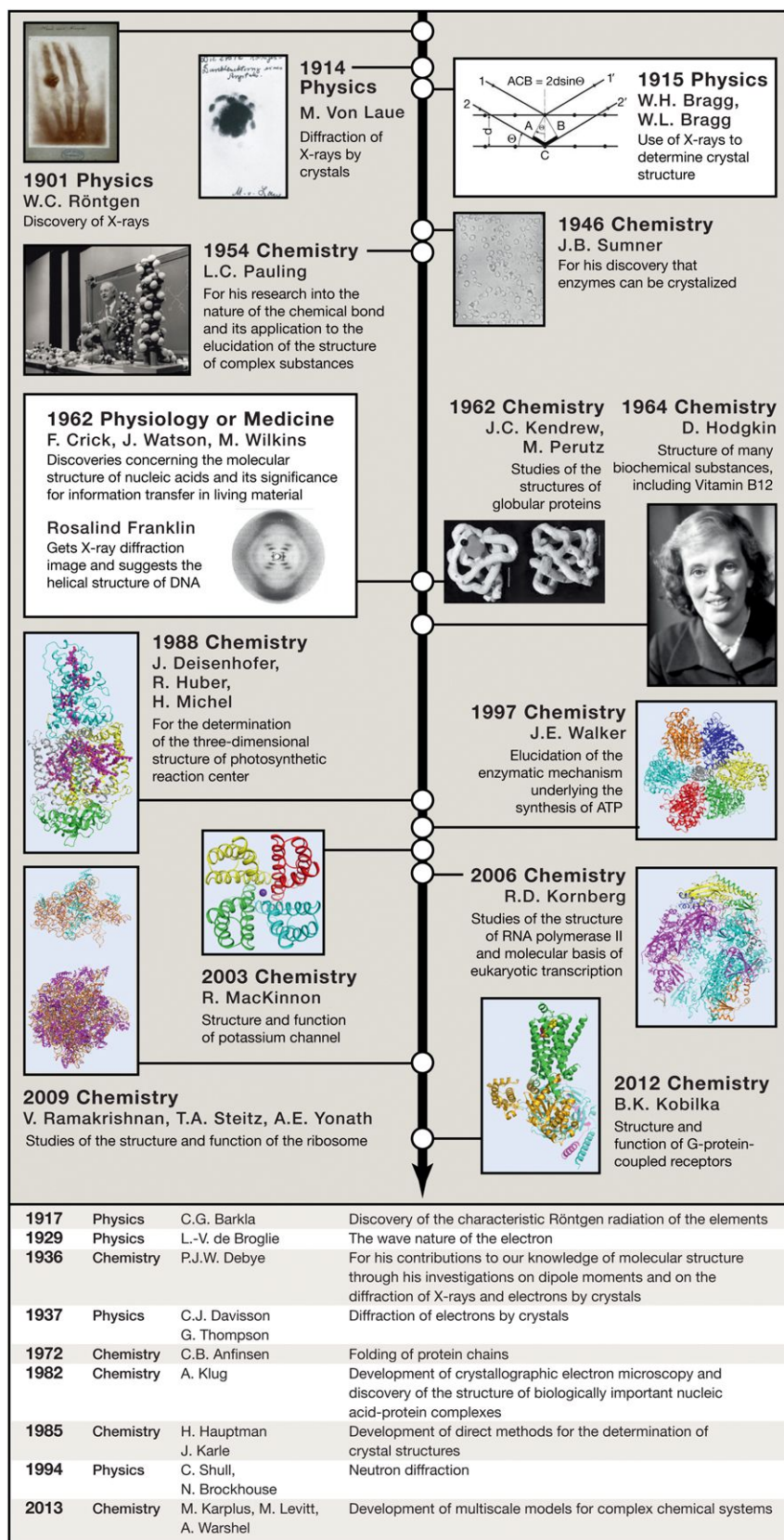


Figure 1. The History of X-Ray Crystallography in the Eyes of Nobel Prizes

Major achievements in the development and application of X-ray crystallography have been recognized by at least 14 Nobel Prizes. The first Nobel Prize in physics was awarded to Roentgen in 1901 for his discovery of X-rays. The next two Nobel Prizes in 1914 and 1915 were given to Laue for his discovery of X-ray diffraction by crystals and to the Bragg father and son for the use of X-rays to determine crystal structure. Sumner was awarded a Chemistry Prize in 1946 for crystallization of the enzyme urease. Pauling won a Chemistry Prize in 1954 for his research into the nature of chemical bond and its application in structure determination. The year 1962 was quite special, with the Chemistry Prize awarded to Kendrew and Perutz for their pioneering work in protein structure determination, and the Physiology or Medicine Prize bestowed on Crick, Watson, and Wilkins for their contribution in the discovery of DNA structure. Hodgkin was awarded a Chemistry Prize in 1964 for structural elucidation of many biochemical substances, including vitamin B12. The next six Nobel Prizes were awarded to macromolecular crystallographers: Deisenhofer, Huber, and Michel in 1988 for the structure of bacterial photosynthetic reaction center; Walker in 1997 for the structure of F1-ATPase; MacKinnon in 2003 for potassium channels; Kornberg in 2006 for the structure of RNA polymerases; Ramakrishnan, Steitz, and Yonath in 2009 for the structure of ribosome; and Kobilka in 2012 for the structure of GPCR. Listed in the lower left corner are nine Nobel Prizes that are closely related to X-ray crystallography.

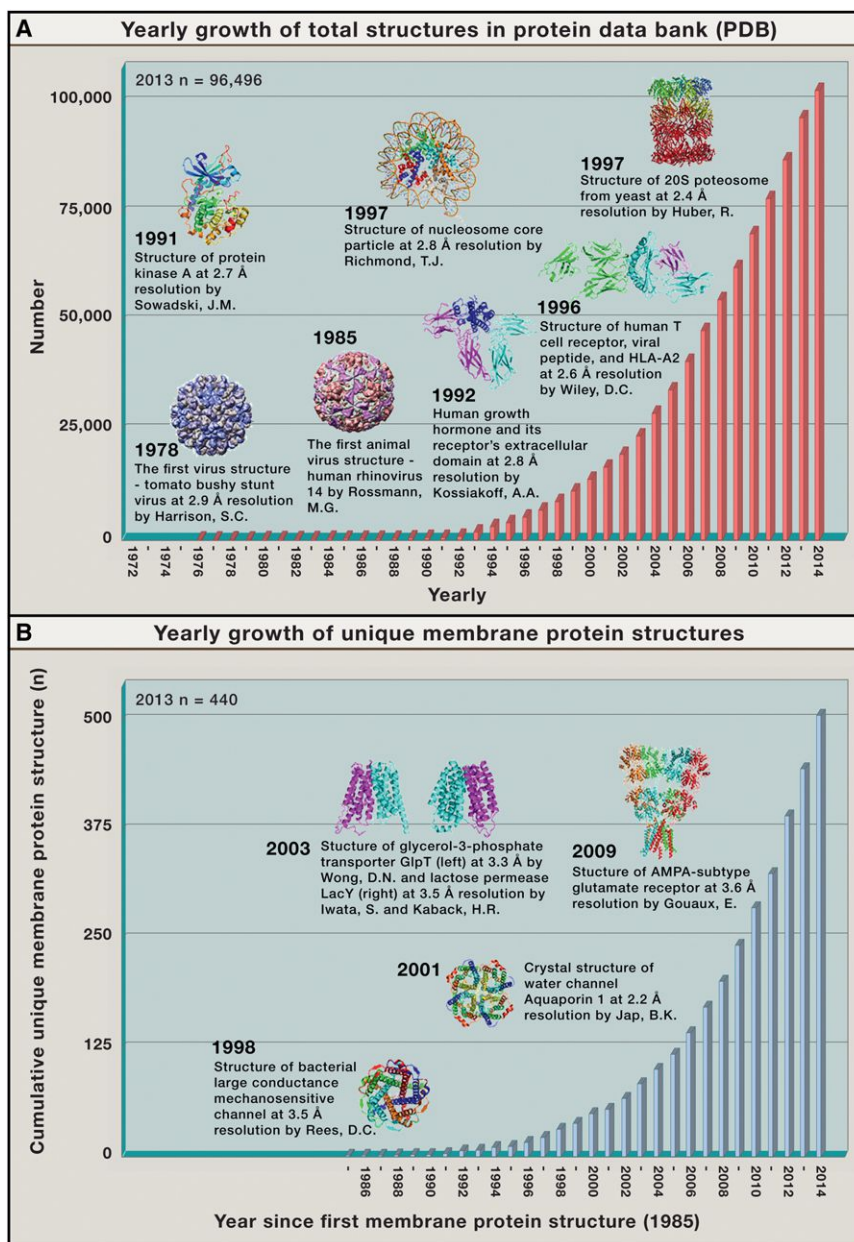


Figure 2. Entries in the Protein Data Bank Have Enjoyed Near-Exponential Growth in the Past 30 Years

(A) The total number of entries of PDB. The PDB was established in 1971 at the Brookhaven National Laboratory with only seven entries. The total number of entries grew to 69 in 1980, 507 in 1990, 13,597 in 2000, and 70,039 in 2010. As of August 26, 2014, there were 102,863 total entries in the PDB, of which 88.7% were determined by X-ray crystallography, 10.3% by NMR, and 0.8% by EM. Some of the representative X-ray structures are indicated in the chart.

(B) The total number of unique membrane protein structures in PDB. In sharp contrast to their heavy presence, structures of membrane proteins only account for just over 1% of all entries in the PDB, with 1,520 total entries and 499 unique structures as of August 31, 2014 (<http://blanco.biomol.uci.edu/mpstruc/>).

technology in the late 1970s—raised the question of how DNA sequences could be specifically recognized by transcription factors. Seeking an answer to this question became an important theme for structural biology in the 1980s (for more details, please see Cramer [2014] in this issue of *Cell*). Many DNA-binding motifs were identified, and a number of crystal structures of protein-DNA complexes were elucidated (Aggarwal et al., 1988; Anderson et al., 1987; Otwinowski et al., 1988; Wolberger et al., 1988). There were many other exciting developments in the 1980s. Michael Rossmann's group determined the first atomic resolution crystal structure of a human common cold virus (Rossmann et al., 1985) (Figure 2A). These early findings were followed up with a barrage of structural information on viruses, including mengo virus (Luo et al., 1987) and foot-and-mouth disease virus (Acharya et al., 1989). Due to impending health threats, disease-causing viruses continue to

attract ample attention from the structural biology community. Insights gained from these virus structures have, in turn, revolutionized our concepts of virology and helped vaccine design and drug discovery. The 1980s also marks the beginning of structural biology on integral membrane proteins, with determination of atomic resolution X-ray structures of the bacterial photosynthetic reaction center (Allen et al., 1987; Deisenhofer et al., 1985). Compared to the 1980s, paces of structural biology were considerably faster in the 1990s. The total number of PDB entries increased from less than 400 at the beginning of the year 1990 to more than 13,000 by the end of the year 2000 (Figure 2A). Cellular signal transduction was the principal focus of biological investigation in the 1990s; consequently, the central theme of structural biology during this period was mechanistic understanding of cell

for the study of protein folding and thermodynamics (Baase et al., 2010). In addition to the study of enzymes, viruses became hotly pursued in the 1960s and 1970s. Following his pioneering research into the tomato bushy stunt virus (Harrison, 1969; Harrison and Jack, 1975), Stephen Harrison and colleagues reported the first virus structure at an atomic resolution of 2.9 Å, revealing 180 copies of the capsid protein arranged in an icosahedral particle (Harrison et al., 1978) (Figure 2A). Aaron Klug and colleagues solved the X-ray structure of the TMV protein disk at 2.8 Å resolution (Bloomer et al., 1978) and obtained a low-resolution structure of nucleosome core particle (Finch et al., 1977; Richmond et al., 1984).

Two exciting developments in biology—discovery of gene regulation in the 1960s and emergence of recombinant DNA

attract ample attention from the structural biology community. Insights gained from these virus structures have, in turn, revolutionized our concepts of virology and helped vaccine design and drug discovery. The 1980s also marks the beginning of structural biology on integral membrane proteins, with determination of atomic resolution X-ray structures of the bacterial photosynthetic reaction center (Allen et al., 1987; Deisenhofer et al., 1985).

Compared to the 1980s, paces of structural biology were considerably faster in the 1990s. The total number of PDB entries increased from less than 400 at the beginning of the year 1990 to more than 13,000 by the end of the year 2000 (Figure 2A). Cellular signal transduction was the principal focus of biological investigation in the 1990s; consequently, the central theme of structural biology during this period was mechanistic understanding of cell

signaling. Crystal structures were elucidated for numerous signaling motifs, exemplified by the phosphotyrosine recognition SH2 domain (Waksman et al., 1992), the phosphoinositol-binding PH domain (Ferguson et al., 1995), the proline-rich sequence binding SH3 domain (Musacchio et al., 1992, 1994), and the carboxyl-terminal peptide-binding PDZ domain (Doyle et al., 1996). The modular nature of these motifs and characterization of their binding specificities constitute an important basis for the development of synthetic biology and chemical biology.

Signal transduction often begins in the extracellular space with ligand binding to its cell-surface receptor. Crystal structure of the complex between human growth hormone and the extracellular domain of its receptor revealed a single hormone molecule recognized by two molecules of receptor (de Vos et al., 1992) (Figure 2A). Some of the other early cocrystal structures include the extracellular domain of TNF receptor bound to TNF- β (Banner et al., 1993), the extracellular fragment of interferon- γ receptor bound to interferon- γ (Walter et al., 1995), and domain 2 of the Flt-1 receptor bound to VEGF (Wiesmann et al., 1997). These ligand-receptor structures give rise to the principle that receptor clustering induced by ligand binding serves as a platform for signaling, which usually involves phosphorylation by receptor kinases. Crystal structure of the catalytic subunit of protein kinase A (PKA)—the first for a protein kinase—revealed an amino-terminal lobe rich in β strands and a carboxyl-terminal lobe with mostly α helices (Knighton et al., 1991) (Figure 2A). The kinase fold observed in PKA was soon confirmed by dozens of crystal structures of other important protein kinases in the 1990s. Understanding of immune signaling by the T cell receptor (TCR) is markedly enhanced by structures of TCR and their complexes with MHC-peptide antigens (Garboczi et al., 1996; Garcia et al., 1996) (Figure 2A).

Structural biology of membrane proteins enjoyed rapid development in the 1990s and beyond. As an extension of X-ray crystallography, electron crystallography was successfully applied to determine the structure of membrane proteins that formed two-dimensional crystals at low resolution in the 1980s and near-atomic resolution in the 1990s. Following two decades of method development on electron crystallography (Subramaniam et al., 2002), Richard Henderson and colleagues succeeded in the elucidation of a 3.5 Å resolution structure of bacteriorhodopsin using electron diffraction data (Grigorieff et al., 1996). Kühlbrandt and colleagues generated a 3.4 Å resolution structure of a plant light-harvesting complex (Kühlbrandt et al., 1994). A striking 1.9 Å resolution was achieved for two-dimensional aquaporin-0 crystals, which allowed clear visualization of lipid-protein interactions (Gonen et al., 2005).

X-ray structure of the mitochondrial F1-ATPase at 2.8 Å resolution revealed distinct conformations for the three catalytic β subunits (Abrahams et al., 1994). Structural analysis of a bacterial potassium channel—the first structure of a recombinant integral membrane protein—provided insights into potassium conduction and selectivity (Doyle et al., 1998). This work was followed up with systematic structural studies of potassium channels, which offer molecular explanations for gating of the ion-conducting pore (MacKinnon, 2003). X-ray structure of the bacterial large-conductance mechanosensitive channel (MscL) revealed a homopentameric assembly and served as a founda-

tion for understanding other mechanosensitive channels (Chang et al., 1998) (Figure 2B). Structural analysis of the water channel aquaporin-1 identified a tetrameric assembly with water molecules localized along a selectivity filter (Figure 2B) (Sui et al., 2001). Crystal structure of the AMPA-type GluA2 receptor revealed a 2-fold symmetry in the extracellular domain and a 4-fold symmetry in the membrane-spanning ion channel domain (Sobolevsky et al., 2009). The first crystal structures on recombinant membrane transporters were elucidated for the lactose permease LacY (Abramson et al., 2003) and the glycerol-3-phosphate transporter GlpT (Huang et al., 2003), defining a conserved fold for the major facilitator superfamily (MFS) of secondary active transporters (Figure 2B). Crystal structure of the human glucose transporter GLUT1—the first eukaryotic MFS structure—allows rationalization of disease-derived mutations (Deng et al., 2014). Following an early X-ray structure on bovine rhodopsin (Palczewski et al., 2000), crystal structures of the recombinant G-protein-coupled receptor (GPCR) were reported on β 2 adrenergic receptor (Cherezov et al., 2007; Rasmussen et al., 2007). Structure of an agonist-bound β 2 adrenergic receptor in complex with a nucleotide-free Gs heterotrimer provided a preliminary answer to how ligand binding to GPCR triggers the activation of G protein in the cytoplasm (Rasmussen et al., 2011).

Improvement of hardware and software in X-ray crystallography has greatly accelerated the pace of discovery and emboldened structural biologists to attack challenging research projects. Structural elucidation of the proteasomal 20S particles revealed the architecture and identified the proteolytic mechanisms (Groll et al., 1997; Löwe et al., 1995; Seemüller et al., 1995) (Figure 2A). Crystal structure of the GroES-GroEL complex markedly improved our understanding of chaperone-assisted protein folding (Xu et al., 1997b), whereas structural analysis of the nucleosome core particle shows in atomic detail how 146 base pairs of DNA are assembled around the histone octamer (Luger et al., 1997) (Figure 2A). Crystal structure of the large ribosomal subunit, including 2833 RNA nucleotides and 27 proteins, was determined at an atomic resolution of 2.4 Å (Ban et al., 2000) and expanded our protein-RNA recognition database as of the year 2000 by a factor of 6-fold. Since the early 1990s, biologically important megacomplexes and macromolecular assemblies have represented increasingly attractive targets for structural biologists. X-ray structures of the eukaryotic exosomes have revealed important insights into the degradation and processing of cellular RNA (Bonneau et al., 2009; Liu et al., 2006; Makino et al., 2013; Wasmuth et al., 2014). Preliminary structural analysis of the U1, U4, and U6 snRNPs, three subcomplexes of the eukaryotic spliceosome, provided mechanistic insights into mRNA splicing (Leung et al., 2011; Montemayor et al., 2014; Pomeranz Krummel et al., 2009; Zhou et al., 2014). Crystal structure of the human COP9 signalosome revealed molecular architecture of the eight-subunit complex (Lingaraju et al., 2014). The combination of X-ray crystallography with single-particle cryo-EM analysis has been routinely used to provide accurate information on megacomplexes, as exemplified by structure of the yeast 26S proteasome (Beck et al., 2012).

The vast amount of information in the PDB serves as the structural basis for understanding biology and innovating medicine.

X-ray crystallography as a tool has played a dominant role in the past five decades in deciphering the molecular mechanisms of virtually all biological processes. It is truly a mission impossible to attempt coverage of, even if just glossing over, all major achievements of structural biology. Because this narrative is targeted to a general readership, the goal is to showcase the spirit of structural biology through brief description of representative examples—how X-ray crystallography has helped transform the understanding of biology. Therefore, I will mainly comment on two areas of biological research that have been galvanized by X-ray crystallography: kinase regulation and membrane protein structure-function relationship (for the areas of chromatin and transcription, please see Cramer [2014] in this issue of *Cell*). Within each research area, the focus is not to provide a comprehensive review but to discuss pioneering structural investigation and select follow-up studies.

Kinases and Anticancer Drugs

cAMP-Dependent PKA

Reversible protein phosphorylation was discovered on glycogen phosphorylase in the 1950s by Edwin Krebs, Edward Fischer, Earl Sutherland, and Wosilait (Fischer and Krebs, 1955; Sutherland and Wosilait, 1955). The first essential step in glycogen metabolism—its phosphorylation—is mediated by glycogen phosphorylase, whose activation depends on phosphorylase kinase. The conversion of inactive phosphorylase kinase to its activated form is mediated by phosphorylase kinase kinase, also known as cyclic AMP (cAMP)-dependent protein kinase, or PKA. The catalytic subunit of PKA in its free form is catalytically active; however, two molecules of the catalytic subunits are constitutively bound and inhibited by a homodimer of regulatory subunits. Binding of the second messenger cAMP to the regulatory subunits results in the dissociation and hence activation of the PKA catalytic subunits.

Human genome is predicted to encode 518 putative protein kinases, which, on the basis of substrate specificity, are classified into 90 tyrosine kinases (PTKs) and 428 serine/threonine kinases (PSKs). These kinases play an essential role in virtually all cellular processes. The very first atomic view on any protein kinase was provided in 1991 by the crystal structure of the catalytic subunit of PKA (Knighton et al., 1991). This structure reveals a general architecture of two lobes, with the substrate-binding site and the ATP-binding pocket both located between the two lobes (Figure 3A). An activation helix (α C), which contains a catalytic triad residue, and an activation loop, whose phosphorylation results in the activation of the kinase, are positioned next to the hinge between the two lobes. These structural features become the hallmarks of nearly all protein kinases.

Structures of the catalytic subunit of PKA explain substrate specificity and define conformational flexibility (Knighton et al., 1991). Structural analysis of the regulatory subunit reveals mechanism of cAMP recognition (Su et al., 1995). Subsequent structural analysis of PKA with both catalytic and regulatory subunits provides a molecular explanation for the inhibition of PKA and identifies a model for cAMP-mediated activation through binding to the regulatory subunit (Figure 3B) (Kim et al., 2005; Wu et al., 2007). A 2.3-Å X-ray structure of the intact, tetrameric PKA holoenzyme yields insights into allosteric regulation and has

ramifications for understanding the regulation of other conserved kinases (Figure 3C) (Zhang et al., 2012).

Cyclin-Dependent Kinases

The temporal-spatial activities of cyclin-dependent kinases (CDKs) determine the fate of the cell cycle and thus are subject to complex regulation. Unlike most kinases, which are active in their free forms, free CDKs are inactive and only attain a basal level of catalytic activity upon association with cyclin. The basal activity can be greatly stimulated by phosphorylation of the activation loop (also known as T loop). The Cip/Kip family of cell-cycle inhibitors, exemplified by p27, potently inhibits the kinase activity of the assembled cyclin-CDK complexes, whereas members of the INK4 inhibitor family can bind the inactive, free CDK, thus preventing its activation. The INK4 inhibitors, exemplified by p16^{INK4a}, also bind and inhibit active cyclin-CDK complexes.

Crystal structure of free CDK2 reveals a misaligned conformation for the activation helix (PSTAIRE helix) and the T loop, providing an immediate explanation to the lack of catalytic activity for the free kinase (De Bondt et al., 1993). Binding by cyclin A induces large conformational changes in the PSTAIRE helix and T loop, resulting in correct alignment of the active site residues and opening of the catalytic cleft (Jeffrey et al., 1995). Phosphorylation of CDK2 in the T loop results in the reorganization of the local structural elements, priming it for substrate binding (Russo et al., 1996b). Remarkably, the inhibitory domain of p27^{Kip1} adopts an extended conformation to interact with both cyclin A and CDK2, with an invariant residue inserting into the catalytic cleft of CDK2 and displacing ATP (Russo et al., 1996a). By contrast, the INK4 inhibitors p16^{INK4a} and p19^{INK4d} bind close to the ATP-binding site of CDK6 and cause structural changes in the catalytic cleft to negatively affect the kinase activity (Brotherton et al., 1998; Russo et al., 1998). The binding site for p16^{INK4a} or p19^{INK4d} is opposite of where cyclin binds, explaining why the INK4 inhibitors can bind both free and cyclin-bound CDKs.

Advent of structural information on CDKs greatly facilitated the screening and design of specific small-molecule inhibitors that block their kinase activities. These inhibitors, mostly ATP analogs, effectively stopped cancer cell growth in petri dish and in animal models. The outcome of anticancer clinical trials thus far has been less than desirable, with notable toxicity and marginal efficacy. For example, flavopiridol exhibited multifaceted, antiproliferative effects in preclinical research but had disappointing performance in clinical trials with narrow therapeutic window and off-target effects. The more selective, second generation CDK inhibitor dinaciclib has entered phase III clinical trial. A notable advance is the development of palbociclib, a highly selective inhibitor of CDK4 and CDK6, for treatment of the HER2-/ER+ breast cancer. In addition, the problem of cross-reactivity by ATP analogs can be effectively dealt with by the development of inhibitors that target sites other than the ATP-binding pocket (Abate et al., 2013).

BCR-ABL, EGFR, and Other Kinase Targets

BCR-ABL is the fusion product of chromosome translocation—a portion of the *Bcr* gene on chromosome 22 is fused to the *Abl* tyrosine kinase gene on chromosome 9 in the Philadelphia chromosome. BCR-ABL is constitutively activated and drives the development of chronic myelogenous leukemia (CML). Perhaps

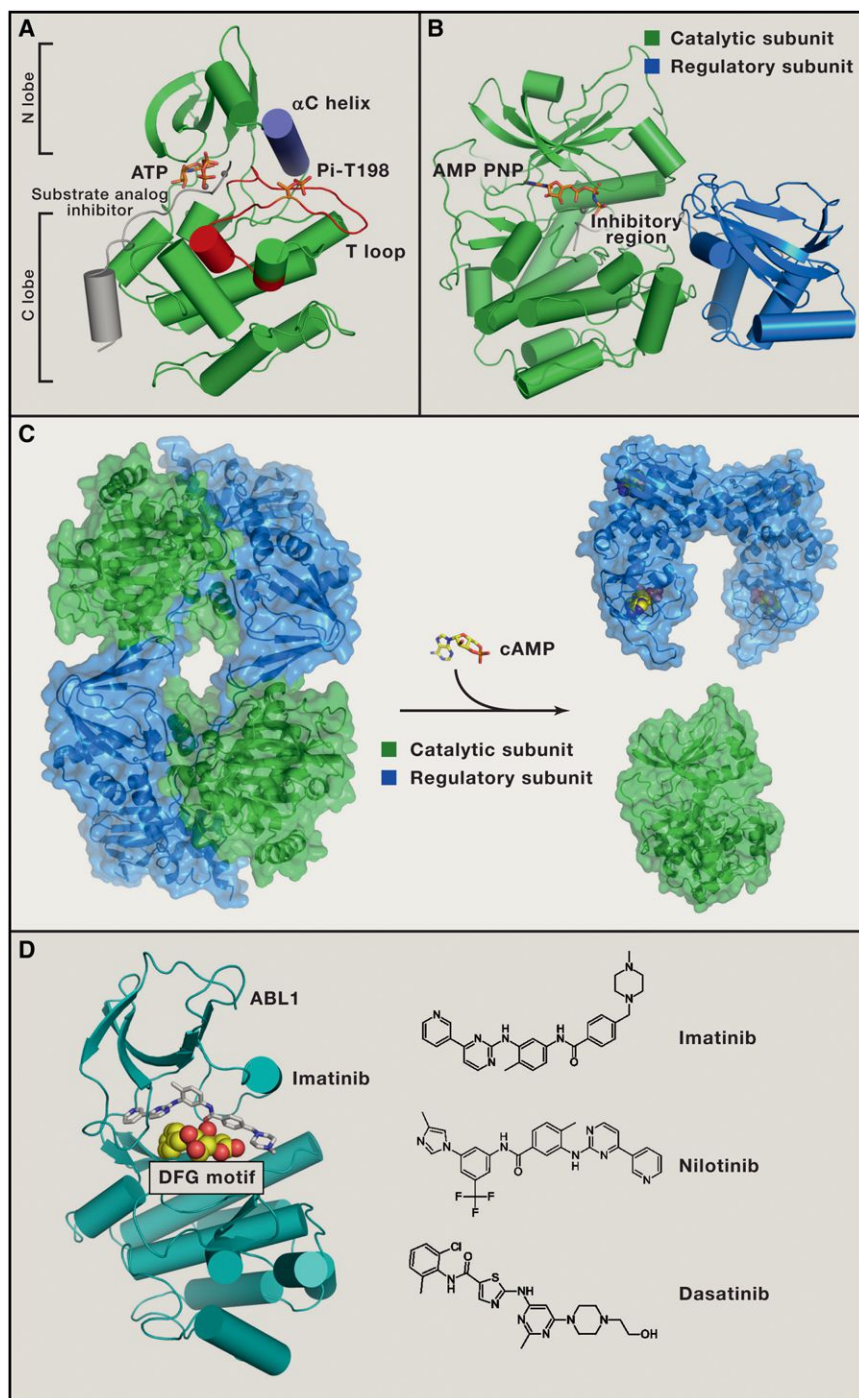


Figure 3. Kinase Structure, Regulation, and Inhibitors

(A) Crystal structure of the catalytic subunit of protein kinase A (PKA, PDB code 1ATP, colored green). The activation helix (α C helix) in the N lobe is colored purple, and the activation loop (T loop) is highlighted in red. ATP and phosphothreonine 198 (Thr198) in the T loop are indicated. The substrate analog inhibitor is represented in gray.

(B) Crystal structure of the complex between PKA catalytic subunit and regulatory subunit (PDB code 3FHI). The PKA catalytic and regulatory subunits are colored green and blue, respectively. The ATP analog AMP-PNP is shown in sticks. The inhibitory region of the regulatory subunit is represented in gray.

(C) Mechanism of cAMP-mediated activation of PKA. In the absence of cAMP, two catalytic subunits (green) and two regulatory subunits (blue) assemble into an inactive, tetrameric PKA holoenzyme (PDB code 3NTF). The binding of cAMP to the regulatory subunits causes pronounced conformational changes, leading to dissociation of the catalytic subunits and producing a dimeric cAMP-bound regulatory subunits (PDB code 4MX3) and two free active catalytic subunits (PDB code 1ATP).

(D) Small-molecule inhibitors of BCR-ABL as potent anticancer drugs. Crystal structure of the human ABL1 kinase domain bound to the anti-cancer drug imatinib (PDB code 2HYY) is shown in the left panel. Imatinib is a potent drug targeting CML through binding and inhibition of BCR-ABL. Two other inhibitors, nilotinib and dasatinib, are used to treat CML patients with imatinib-resistant mutations. The three small molecule inhibitors are shown in the right panel.

tures reveal the precise interactions of an imatinib variant with surrounding residues in the ATP-binding pocket of BCR-ABL and explain how patient-derived mutations (such as T315I) inactivate imatinib binding (Figure 3D) (Nagar et al., 2002; Schindler et al., 2000). Imatinib, classified as a type II kinase inhibitor, binds the kinase in its inactive DFG-out conformation. The structural information guided subsequent development of second-generation inhibitors that stabilize these kinase-specific inactive conformations. These inhibitors include ponatinib, which is effective against the T315I mutation, and nilotinib (sometimes known as super gleevec, Figure 3D), which works well against the majority of CML mutations (except T315I).

The structural information also helped design of the type I kinase inhibitors, exemplified by dasatinib (Figure 3D), which bind the kinase in its active DFG-in conformation. These second-generation inhibitors are used to treat relapsed CML patients with imatinib-resistant mutations.

The epidermal growth factor receptor (EGFR) family of receptor tyrosine kinases (RTK) has four members ErbB1–4, each with an extracellular ligand-binding domain, a single transmembrane

the most celebrated, anticancer kinase inhibitor is imatinib, also known as gleevec or STI-571, which targets BCR-ABL (Shah and Sawyers, 2003). Nearly all chronic myelogenous leukemia (CML) patients who took the drug in the first phase I trial in 1998 were responding; imatinib was approved by FDA in 2001. Imatinib also potently inhibits the tyrosine kinases c-kit and PDGFR, and its clinical use has been approved for a few other cancer types associated with c-kit or PDGFR activation. Crystal struc-

segment, and an intracellular kinase domain. The activation mechanism of EGFR has been elegantly revealed by crystallographic discovery and subsequent biochemical analysis (Burgess et al., 2003; Zhang et al., 2006). Growth factor binding induces conformational changes in the extracellular domain, leading to its dimerization with the two ligands on opposite sides of the heterocomplex (Garrett et al., 2002; Ogiso et al., 2002). The EGFR kinase domain exists in an autoinhibited conformation, with key residues Leu834 and Leu837 stabilizing the inactive conformation of the α C helix. Receptor dimerization, induced either by ligand binding or by high protein concentrations during crystallization, facilitates formation of an asymmetric dimer, where the C-lobe of the activator kinase interacts with the N-lobe of the receiver kinase, leading to allosteric activation of the latter (Zhang et al., 2006). The structural observations also provide a satisfying explanation to how mutations of Leu834/Leu837 drive EGFR activation. Aberrant activation of EGFR contributes to the development of a number of malignant cancer types. A few small-molecule inhibitors of EGFR have been used in the clinic to treat cancers, such as gefitinib (popularly known as Iressa) for treating non-small-cell lung cancer (NSCLC). Patients with activating mutations in EGFR respond very well to the treatment of gefitinib (Lynch et al., 2004; Paez et al., 2004). Other small-molecule inhibitors targeting EGFR include lapatinib for breast cancer and erlotinib (or Tarceva) for NSCLC and pancreatic cancer.

Monoclonal antibodies have also been developed to prevent ligand binding to EGFRs or to sabotage dimerization of the extracellular domain. Herceptin, or trastuzumab, which blocks ligand binding by associating with the extracellular domain of ErbB2/HER2 (Cho et al., 2003), proved to be effective in treating ErbB2-overexpressing breast cancer (Recondo et al., 2014). Pertuzumab, on the other hand, prevents receptor dimerization by binding to the extracellular domain of ErbB2/HER2 (Franklin et al., 2004). Cetuximab, or Erbitux, which prevents both ligand binding and receptor dimerization through binding to domain III of the EGFR extracellular region (Li et al., 2005), has been approved for the treatment of metastatic colon cancer, NSCLC, and head and neck cancer. More recently, monoclonal antibody and cytotoxic small molecule have been combined into a single entity, named antibody-drug conjugates, which directly target cancer cells with high-dose chemotherapy. Trastuzumab emtansine (or Kadcyla) is such an antibody-drug conjugate that combines trastuzumab and the cytotoxic agent mertansine; it is approved to treat HER2-positive metastatic breast cancer (Recondo et al., 2014).

Chemical Genetics of Protein Kinases

The structural information not only facilitated drug discovery targeting various malignancies but also gave birth to chemical genetics on kinases. The proto-oncogene c-Src, discovered by J. Michael Bishop and Harold Varmus, exists in normal cells in an inactive conformation. In the crystals of autoinhibited c-Src, the SH3 domain, the SH2 domain, and the kinase domain interact with each other to assemble into a tightly folded assembly in which the phosphorylated Tyr527 binds to the SH2 domain and locks c-Src in an inactive conformation (Sicheri et al., 1997; Xu et al., 1997a). Activation of the c-Src pathway, triggered by dephosphorylation of Tyr527, contributes to multiple cancer

types. On the basis of structural modeling, the ATP-binding site of v-Src was mutated such that the resulting v-Src, but not the wild-type (WT) v-Src, could accept a synthetic ATP analog (Shah et al., 1997). The engineered v-Src displayed similar catalytic efficiency, as well as substrate specificity as the WT v-Src, but allowed direct tracing of v-Src substrates in cells (Shah et al., 1997). This strategy was applied to other Src family members and general protein kinases. The engineered kinases, along with the unique ATP analogs, allowed investigation of important biological questions, such as identification of Cdk1 targets in yeast, selective inhibition of neurotrophin *in vivo*, and discovery of JNK2 as a positive regulator of c-Jun.

Membrane Protein Structure and Function

Membranes compartmentalize cellular processes and enzymatic reactions, and membrane proteins account for 20%–30% of all human proteins. Structures of membrane proteins, however, only account for just over 1% of all entries in the PDB, with 1,520 total entries and 499 unique structures as of August 31, 2014 (Figure 2B). The structural investigation of membrane proteins had been hampered by the technical challenges of poor recombinant expression, insolubility in aqueous solution, and unruly behavior in detergent solubilized forms. In addition, crystals of membrane protein usually diffract X-rays poorly. Consequently, structural biology of membrane proteins lagged behind that for soluble proteins by two decades.

Early effort focused on endogenous membrane proteins, eliminating the hassles of recombinant expression. Crystal structure of a bacterial photosynthetic reaction center at 3 Å resolution—the first atomic-resolution image of any integral membrane protein—reveals the stunning inner workings usually buried within the membrane (Deisenhofer et al., 1985) (Figure 1). X-ray structure of the F1-ATPase from bovine heart mitochondria, determined at 2.8 Å resolution, captured the three catalytic β subunits in distinct conformations and different states of nucleotide binding (Abrahams et al., 1994) (Figure 1). This structural observation lends critical support to the hypothesis that the three catalytic subunits are at different states of the catalytic cycle at any moment, and rotation of the $\alpha 3\beta 3$ subcomplex relative to the γ subunits results in the interconversion of the states.

Roderick McKinnon was among the first to use recombinant system to express membrane proteins—potassium channels—for structural studies (Figure 1). Despite similar properties between potassium (K^+) and sodium (Na^+), K^+ channels are at least four orders of magnitude more permeant for K^+ than for Na^+ . Crystal structure of the K^+ channel from *Streptomyces lividans* (KcsA) at 3.2 Å resolution revealed the first atomic view of an ion channel (Figure 4A) (Doyle et al., 1998). Tetrameric arrangement of the K^+ selectivity filter and structural features of the pore provide explanations to ion selectivity and conduction (Doyle et al., 1998; Roux and MacKinnon, 1999; Zhou et al., 2001). Only two out of four potential K^+ binding sites are occupied in the selectivity filter, allowing energy minimization and optimization of conduction rate (Morais-Cabral et al., 2001). Subsequent structural and biochemical studies on K^+ channels uncovered mechanistic insights into gating of the ion-conducting pore (Jiang et al., 2002, 2003; Long et al., 2005).

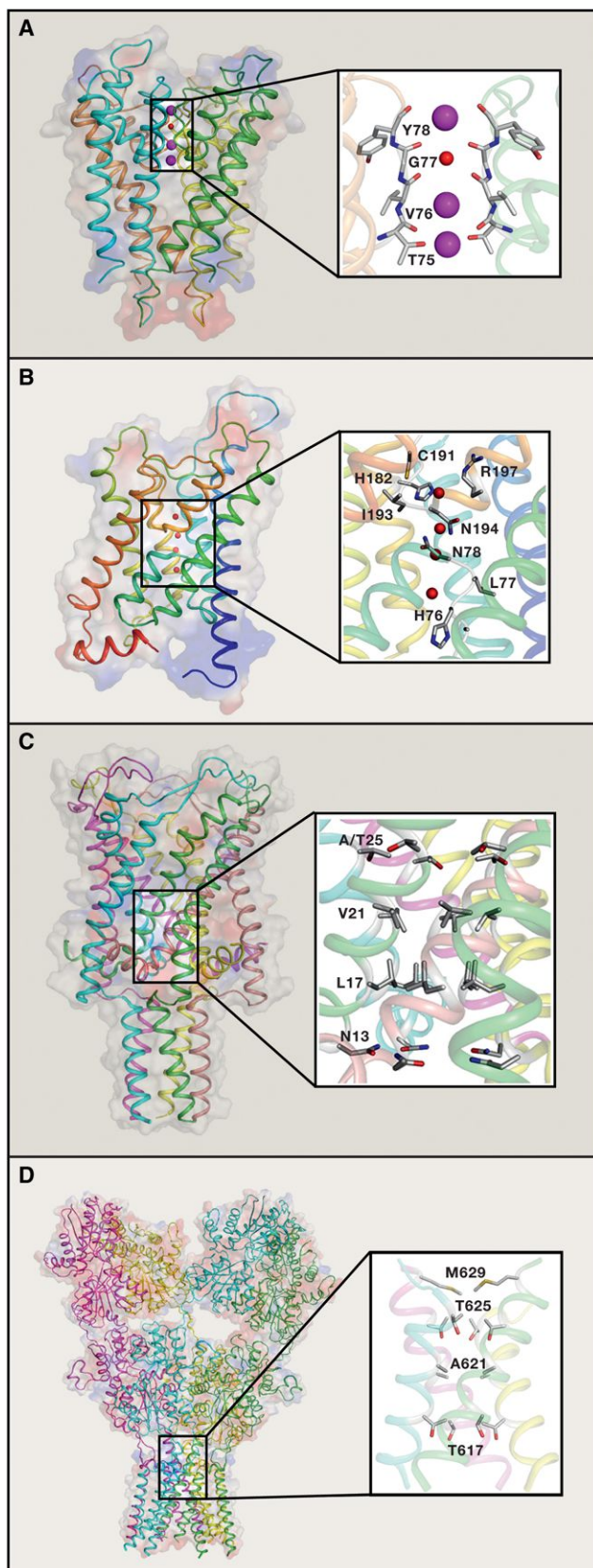


Figure 4. Structures and Mechanisms of Representative Channels

(A) Crystal structure of the potassium channel KcsA (PDB code: 1BL8). KcsA is a homotetramer. Each subunit contains two α helices connected by the pore region, which harbors the selectivity filter. The selectivity filter is shown in a close-up view, with three K^+ ions (magenta spheres) identified. Two K^+ ions are in a rapid equilibrium, with only one of the lower two positions occupied at any instant. One water molecule is shown in red sphere.

(B) Crystal structure of the water channel aquaporin 1 (AQP1) (PDB code: 1J4N). AQP1 is a tetramer with each subunit providing an independent water pore. Four water molecules (red spheres) are located at three hydrophilic nodes along the selectivity filter.

(C) Crystal structure of the mechanosensitive channel MscL (PDB code: 2OAR). MscL is organized as a homopentamer. A water-filled opening from the extracellular side narrows at the cytoplasmic side, where hydrophilic residues may function as the selectivity filter.

(D) Crystal structure of the AMPA-type glutamate receptor GluA2 (PDB code: 3KG2). The GluA2 structure is a homotetramer. In the antagonist-bound structure, the ion channel adopts a closed conformation. Residues forming the narrowest region are indicated in the close-up view.

G-Protein-Coupled Receptors

GPCRs define a large family of seven transmembrane proteins that mediate a wide range of signaling at the plasma membrane. Approximately half of all clinical drugs directly target GPCRs. Working with Robert Lefkowitz, Brian Kobilka cloned and biochemically characterized human α_2 - and β_2 -adrenergic receptors (Kobilka et al., 1987a, 1987b, 1988). X-ray structure of bacteriorhodopsin, which bears homology to mammalian GPCR, was determined in 1997 (Pebay-Peyroula et al., 1997), followed by the structure determination of bovine rhodopsin (Palczewski et al., 2000). Conformation of the seven transmembrane helices (TMs) in bovine rhodopsin differs significantly from that in bacteriorhodopsin. Kobilka and colleagues determined the crystal structure of the human β_2 adrenergic receptor (β_2 AR) at 3.4–3.7 Å resolution (Rasmussen et al., 2007). The relatively poor X-ray diffraction of β_2 AR crystals was successfully mitigated by insertion of T4 lysozyme into the third intracellular loop (Rosebaum et al., 2007). Crystal structure of the resulting β_2 AR bound to a diffusible ligand carazolol was determined at 2.4 Å resolution, revealing extensive interactions of carazolol with residues at the ligand-binding site (Cherezov et al., 2007). Structures of activated and/or agonist-bound, as well as antagonist-bound, GPCRs reveal distinct conformations of the ligand-binding pocket. The most notable ligand-induced conformational change on the cytoplasmic side appears to be an outward movement of the cytoplasmic portion of TM5 and TM6. A wealth of rapidly emerging structures on GPCRs has greatly stimulated the interests of major pharmaceutical companies to improve existing drugs and to screening and design new therapeutic modulators.

The principal biological question on GPCR is how conformational changes triggered by ligand binding result in the activation of G protein. A tentative answer to this question was supplied by the crystal structure of an agonist-bound β_2 AR in complex with a nucleotide-free Gs heterotrimer (Rasmussen et al., 2011) (Figure 1). The most pronounced, agonist-induced conformational change in β_2 AR is a 14 Å outward movement at the cytoplasmic end of TM6 and TM5. The conformational changes induced by the interactions between β_2 AR and Gs are propagated to the nucleotide-binding pocket, presumably facilitating replacement of GDP by GTP. The most unanticipated change is a marked displacement of the α -helical domain of G α_s relative to the Ras-like GTPase domain (Rasmussen et al., 2011).

Membrane Transport Proteins

A living cell constantly needs to uptake nutrients from the environment and to expel metabolites and waste materials. This extremely complex process is mediated by a very large number of membrane transport proteins that can be classified into at least four general types: channels or pores, facilitators, secondary active transporters, and primary active transporters. An online database of membrane transport proteins (www.TCDB.org) contains more than 10,000 unique protein sequences that are classified into over 800 transporter families. Membrane proteins of known three-dimensional structure are cataloged online at <http://blanco.biomol.uci.edu/mpstruc/>. I do not intend to comprehensively cover all known structures of transporters. Rather, I will simply give a few representative examples.

Channels. The aquaporin family of channels, consisting of the water-conducting aquaporins and the glycerol-conducting aquaglyceroporins, plays an essential role in the regulation of cellular osmolarity. Structural analysis of aquaporin-1 (AQP1) revealed a tetrameric assembly and identified four water molecules that are localized at three hydrophilic nodes along an otherwise highly hydrophobic selectivity filter (Figure 4B), supporting rapid water transport (Sui et al., 2001). Subsequent structural analysis of aquaporin Z identified the molecular mechanism for differentiation between water and glycerol by the aquaporin family (Savage et al., 2003). Structure of the glycerol facilitator GlpF revealed an amphipathic selectivity pore that is lined by a number of glycerol molecules in single file (Fu et al., 2000). Structural analysis and molecular dynamics simulation elucidated the molecular mechanism of selective permeability for glycerol (Fu et al., 2000; Tajkhorshid et al., 2002). The aquaporin fold is also observed in other small-molecule transporters such as the pentameric formate channel FocA (Waight et al., 2010; Wang et al., 2009).

Physical forces, in the form of touch, hearing, pressure, and gravity, are primarily sensed by a family of mechanosensitive ion channels, which transduce mechanical strain into an electrochemical response. The X-ray structure of the MscL, determined at 3.5 Å resolution, revealed a homopentamer (Figure 4C) (Chang et al., 1998). In each subunit, a water-filled opening at the extracellular side is followed by a hydrophilic pore that narrows to an occluded region at the cytoplasmic side. This structure serves as a model system for understanding of and comparison with other mechanosensitive channels. Structural analysis of the small-conductance mechanosensitive channel (MscS) shows a heptameric assembly, with the closed-state transmembrane pore connecting to a large chamber in the cytoplasmic side (Bass et al., 2002). Notably, the overall fold in the transmembrane region is different between MscL and MscS, with two TMs in MscL and three TMs in MscS.

Chemical transmitters of excitatory synapses in the central nervous system, exemplified by glutamate, activate receptors on the postsynaptic cells, leading to transmission of signals from one neuron to the next. Ionotropic glutamate receptors (iGluRs) are ligand-gated ion channels. The iGluR family includes AMPA, kainite, and NMDA receptors, which are heterotetrameric or homotetrameric. Crystal structure of the homotetrameric AMPA-type GluA2 receptor bound to a competitive antagonist revealed a closed conformation, with a 2-fold symmetry in the

extracellular domain and a 4-fold symmetry in the membrane-spanning ion channel domain (Figure 4D) (Sobolevsky et al., 2009). Structural analysis of the GluA2 receptor allowed proposition of mechanisms for ion channel activation, desensitization, and inhibition by noncompetitive antagonists. These mechanisms were further investigated by the X-ray structure of GluA2 bound to cone snail toxin (Chen et al., 2014) and in distinct functional states (Dürr et al., 2014). Structure of the NMDA-type, heterotetrameric receptor involving two GluN1 and two GluN2B subunits confirmed some of the proposed general mechanisms for iGluRs and provided additional insights (Karakas and Furukawa, 2014; Lee et al., 2014).

Secondary Active Transporters. Unlike channels, transporters never allow simultaneous substrate access from both sides of the lipid membrane. A prevailing model for the general transport mechanism is known as alternating access (Jardetzky, 1966), which postulates that a transporter must switch between at least two conformations: open to the extracellular side (outward-open) for substrate upload and open to the intracellular side (inward-open) for substrate release, or vice versa. Restrained by a series of conformational switches within each transport cycle, a transporter can only move 200–50,000 substrate molecules per second, considerably slower than that for channels (which sometimes can permeate substrates at near diffusion limit). The ability to transport substrate molecules against their concentration gradients, together with the different conformational states within each transport cycle, have made membrane transporters appealing for structural and mechanistic investigations.

The major facilitator superfamily (MFS) transporters are ubiquitously present in all kingdoms of life and play an important role in numerous cellular processes. The first crystal structures on MFS were reported on the lactose permease LacY from *E. coli* (Abramson et al., 2003) and the glycerol-3-phosphate transporter GlpT (Huang et al., 2003) (Figure 5A). The structures revealed a conserved MFS fold of 2-fold symmetric N-domain and C-domain, each comprising six consecutive TMs. The transport path, as well as the substrate-binding site, is located between the N and C domains. The structural information provides immediate explanation to a large body of biochemical and biological observations, particularly on LacY (Kaback, 2005). Distinct conformational states of MFS, which include the multidrug transporter EmrD (Yin et al., 2006), the L-fucose:H⁺ symporter FucP (Dang et al., 2010), the peptide transporters PepT_{So} and PepT_{St} (Newstead et al., 2011; Solcan et al., 2012), and the D-xylose:H⁺ symporter Xyle (Sun et al., 2012) appear to support the alternating access model. Perhaps the best known and most important MFS family members are the glucose transporters GLUT1–4, which are associated with a number of debilitating diseases. Crystal structure of the full-length human GLUT1—the first eukaryotic MFS structure—allows rationalization of disease-derived mutations (Figure 5B) (Deng et al., 2014).

Crystal structure of the sodium:leucine symporter LeuT revealed a conserved fold of ten TMs grouped into two inverted repeats, each containing five consecutive TMs (Figure 5C) (Yamashita et al., 2005). The first TM in each of the two inverted repeats, TM1 or TM6, is discontinuous and contains a highly conserved unwound segment that is positioned in the transport path. A similar fold was subsequently observed in the

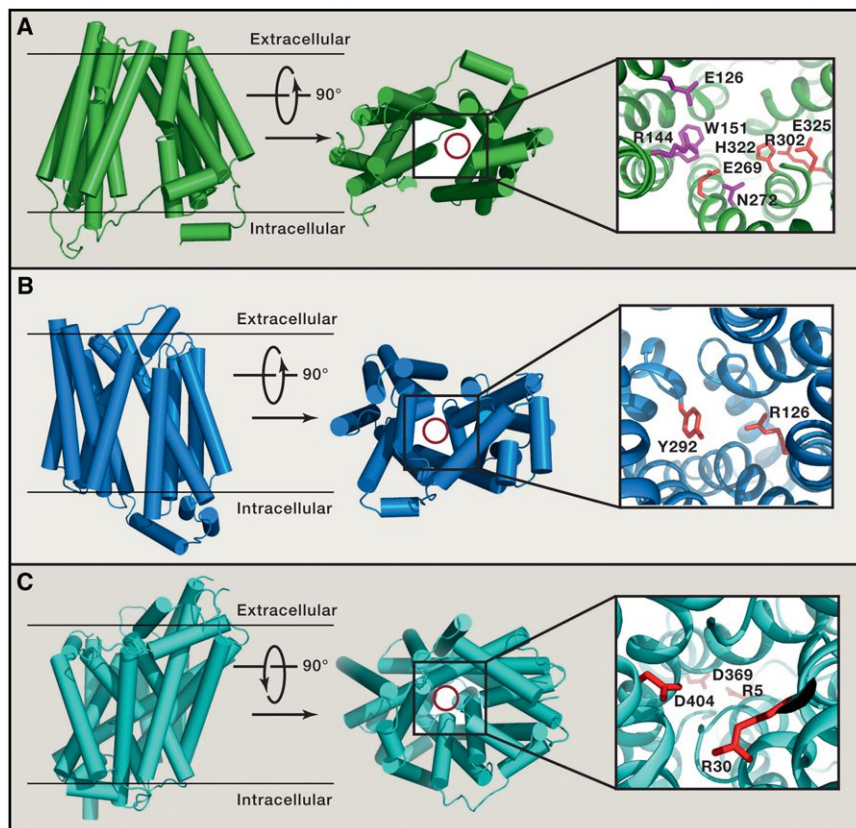


Figure 5. Structures and Mechanisms of Representative Secondary Active Transporters

(A) Crystal structure of the *E. coli* lactose permease LacY. The overall structure of LacY is shown in two perpendicular views (left and middle). The transport path is highlighted by a red circle in the middle panel, and its close-up view is shown in the right panel. Residues involved in substrate binding and proton coupling are colored magenta and red, respectively.

(B) Crystal structure of the human glucose transporter GLUT1. Two amino acids forming the extracellular gate are labeled in the close-up view.

(C) Crystal structure of the sodium symporter LeuT. Arg30 and Asp404 are thought to play a key role at the extracellular gate, whereas Arg5 and Asp369 control the intracellular gate. These four amino acids are indicated in the close-up view.

nucleobase:cation symporter-1 family Mhp1 (Weyand et al., 2008) and the solute sodium symporter family vSGLT (Faham et al., 2008). Remarkably, despite lack of sequence or functional conservation, the same general LeuT fold has been recognized in several membrane transporter families, exemplified by the glycine:betaine transporter BetP (Ressl et al., 2009), and amino acid antiporters AdiC and GadC (Fang et al., 2009; Gao et al., 2009, 2010; Ma et al., 2012). These unanticipated structural findings led to reclassification of membrane transporters, with nine LeuT-fold families now grouped together to constitute the APC superfamily (Saier et al., 2009).

Primary Active Transporters. Primary active transporters, exemplified by the sarco/endoplasmic Ca^{2+} -ATPase (SERCA) and the ATP binding cassette (ABC) transporters, exploit the energy of ATP binding and hydrolysis for substrate transport, usually against concentration gradient of the substrate molecules. The 2.6 Å resolution crystal structure of SERCA1a revealed a complex architecture, with two calcium ions bound in the membrane spanning region (Figure 6A) (Toyoshima et al., 2000). Structural comparison with the enzyme in the absence of calcium suggests large domain movements during active transport, which was confirmed by the structure of SERCA1a in a calcium-free state (Toyoshima and Nomura, 2002). The active transport of calcium ion by this ATPase pump involves several distinct conformational states. These conformational states were visualized by a number of related crystal structures of SERCA, including that bound to ATP or an ATP analog (Olesen et al., 2007; Olesen et al., 2004; Sørensen et al., 2004; Toyoshima

and Mizutani, 2004), ADP (Sørensen et al., 2004), and a phosphate analog (Olesen et al., 2007; Toyoshima et al., 2004). Together, these structures allow mechanistic understanding of the complete cycle of calcium transport driven by ATP binding and hydrolysis.

An ABC transporter contains at least four subunits: two transmembrane domains (TMDs) and two ABCs (or nucleotide binding domains, NBDs) located in the cytoplasm. The first crystal structure

of the ABC transporter—that of the tetrameric BtuCD complex at 3.2 Å resolution—revealed a central translocation path that is formed between two BtuC subunits within the membrane and closed to the cytoplasmic side by a gate region (Figure 6B) (Locher et al., 2002). This structure serves as a framework for all subsequent structural investigations and mechanistic understanding. Crystal structures, including a bacterial multidrug transporter Sav1866 (Dawson and Locher, 2006), a putative metal-chelate-type transporter (Pinkett et al., 2007), a maltose transporter (Oldham et al., 2007), BtuCD bound to the periplasmic-binding protein BtuF (Hvorup et al., 2007), a bacterial lipid flippase MsbA (Ward et al., 2007), a molybdate/tungstate transporter ModBC (Gerber et al., 2008), and a methionine transporter (Kadaba et al., 2008), captured different conformational and nucleotide-bound states. These snapshots allow proposition of a complete transport cycle for an ABC transporter. The transport cycle is elegantly shown for the maltose transporter MalFGK2, with a number of conformational states observed at atomic details (Chen et al., 2013; Khare et al., 2009; Oldham and Chen, 2011a, 2011b; Oldham et al., 2007).

The energy-coupling factor (ECF) family of membrane transporters shares a similar organization with the ABC transporters with two NBDs and two TMDs. In contrast to the ABC transporters, the two NBDs of an ECF transporter are encoded by two different ATPase genes, and the two transmembrane domains have contrasting roles, with one specific for substrate binding (S component) and the other for energy transduction (T component). Structures of the heterotetrameric ECF

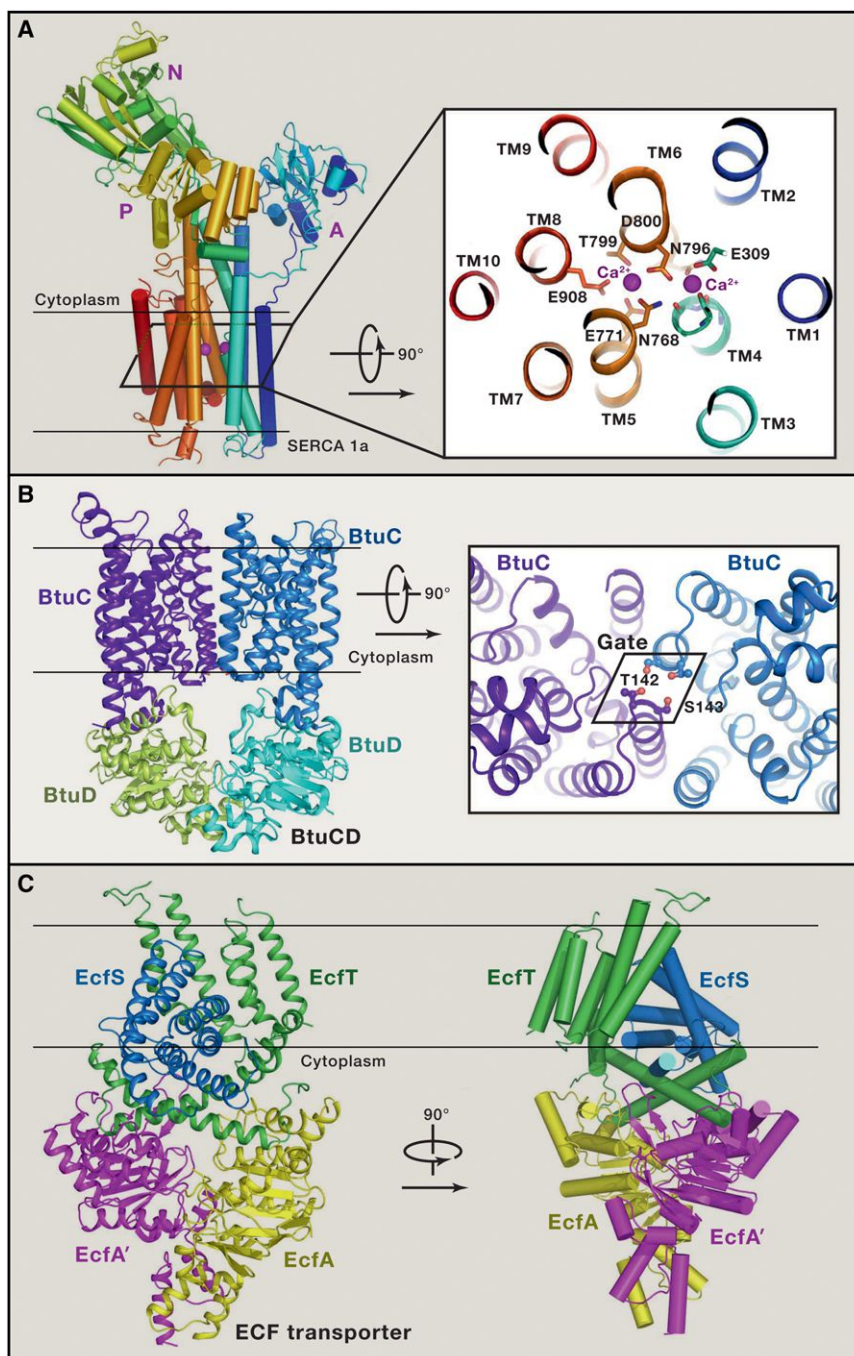


Figure 6. Structures and Mechanisms of Representative Primary Active Transporters

(A) Crystal structure of the calcium ATPase SERCA1a. The structure of SERCA1a is displayed in a rainbow-colored cartoon representation with its amino-terminus in blue and carboxyl-terminus in red (left). Three cytoplasmic domains, A, N, and P, are labeled, and the transmembrane domain consists of ten TMs. The calcium binding sites are shown in a close-up view (right). Two calcium ions are represented in magenta spheres.

(B) Crystal structure of the ABC transporter BtuCD. BtuCD consists of four subunits, two membrane-spanning subunits BtuC (purple and blue), and two ABC subunits BtuD (cyan and green). The transport path is closed to the cytoplasm by a gate region, with two residues Thr142 and Ser143 playing a key role.

(C) Crystal structure of a representative ECF transporter. The structure of a tetrameric ECF transporter is shown in two perpendicular views. The TMs of the substrate-binding S component (EcfS) are nearly parallel to the lipid membrane and perpendicular to those of the energy-transducing T component (EcfT). The structural organization revives the carrier model for membrane transporters.

bilayer, is a universally conserved signaling mechanism from bacteria to human beings (Brown et al., 2000). The most remarkable examples of RIP include cleavage of the cholesterol-controlling transcription factor SREBP by the metalloprotease S2P, the cleavage of the growth factor Spitz by the serine protease rhomboid, and the cleavage of amyloid precursor protein by the aspartate protease presenilin (which is the catalytic component of the γ -secretase). Unlike proteases in aqueous environment, the hydrophilic active site of an intramembrane protease must be accessible to water molecules, which are required for proteolysis, and substrate proteins within the hydrophobic lipid membrane. Structural information is now available on the bacterial homologs of the three classes of intramembrane proteases, including the rhomboid homolog GlpG from *E. coli* (Ben-Shem et al., 2007; Lemieux et al.,

transporters revealed a striking organization—the six TMs of the S component are placed roughly parallel to the lipid membrane and perpendicular to the TMs of the T component (Figure 6C) (Wang et al., 2013; Xu et al., 2013). These structures, together with results of MD simulation, strongly support the carrier model of substrate transport (Widdas, 1952), which differs from the alternating access model.

Intramembrane Proteases. Regulated intramembrane proteolysis (RIP), which requires cleavage of a membrane-embedded substrate protein by an intramembrane protease within the lipid

2007; Wang et al., 2006; Wu et al., 2006), the archaeal S2P homolog (Feng et al., 2007), and the presenilin homolog PSH (Li et al., 2013). In all cases, the active site has constant access to water molecules through a funnel that opens either to the extracellular side or the cytoplasmic side. This crystallographic finding is elegantly simple yet completely unanticipated in the absence of structural information. The active site is shielded from the hydrophobic membrane bilayer by one or two gating TMs, which undergo lateral rotation or movement to allow substrate entry. For GlpG, the gating helix TM5 had been captured in fully

open, partially open, and closed conformations. The recent cryo-EM structure of human γ -secretase (Lu et al., 2014), along with the X-ray structure of its putative substrate-recruiting component nicastrin (Xie et al., 2014), revealed tantalizing clues about functional mechanisms of γ -secretase.

Method Development

Protein Expression and Purification

Prior to 1970s, all proteins employed for X-ray crystallography were derived from endogenous sources—mostly animal tissues and organs—and were biochemically purified. These proteins, exemplified by myoglobin and lysozyme, are thermodynamically stable and contain few flexible surface sequences that may hinder crystallization. In late 1970s, discovery of the restriction enzymes greatly facilitated cloning of target cDNA into plasmids, allowing recombinant expression in *E. coli*. Unfortunately, many recombinant proteins, especially those containing multiple domains and/or flexible surface sequences, defied crystallization. Finding a crystallizable protein domain or fragment frequently required multiple trials of protein engineering, each involving different boundaries or mutations for the target protein. Limited proteolysis has been used to identify stable core domains that are amenable for crystallization. One of the most remarkable examples of limited proteolysis was the identification of sequence boundaries for the DNA-binding domain of p53 (Pavletich et al., 1993).

Recombinant coexpression of two or more proteins became a popular strategy in the 1990s. This strategy allowed convenient assembly of multicomponent protein complex. Although bacterial expression is fast and inexpensive, eukaryotic proteins are frequently misfolded or aggregated when expressed in bacteria, likely due to differences in folding environment. Eukaryotic recombinant expression systems, involving yeast, baculovirus-infected insect cells, and mammalian cells, play an increasingly more important role (Assenberg et al., 2013). For mammalian cell expression, both stable cell lines and transient expression through plasmids have succeeded in production of large quantities of materials for crystallization. Stable CHO cells allowed expression and purification of milligram quantity of γ -secretase, a four-component membrane protein complex (Alattia et al., 2013).

Crystallization

Until early 1990s, approximately half of all crystals were generated with ammonium sulfate as the major precipitant in the crystallization buffer. In the year 2013, however, none of the 55 crystal structures reported in the journal *Nature* relied on ammonium sulfate as a crystallization agent. This dramatic change likely reflects the proposition that ammonium sulfate may work better for small protein of greater thermal stability. Perhaps more importantly, the reagents and tools for crystallization have steadily improved in the last two decades, including numerous commercial reagent kits, robotic arms, and development of special methodology for membrane protein crystallization (McPherson and Gavira, 2014). The lipidic cubic phase (LCP) has been widely used for the crystallization of membrane proteins (Caffrey, 2009), with fantastic successes for GPCRs (Cherezov et al., 2010). Bicelles were also applied to membrane protein crystallization (Agah and Faham, 2012).

Although prior experience may facilitate crystallization of a new target protein or macromolecular complex, systematic screening of representative crystallization space is almost always recommended. In all cases of difficult proteins, biochemistry and molecular biology are the keys for crystallization—that is, improvement of solution behavior through biochemical characterization and extensive protein engineering. Insertion of a T4 lysozyme into β 2AR proved to be important for improvement of crystal packing and hence X-ray diffraction limit (Rosebaum et al., 2007). Sequence alignment among homologous proteins frequently yields valuable hints on how to improve the chance of crystallization. For example, replacing five nonconserved amino acids with conserved ones led to crystallization of a presenilin homolog (Li et al., 2013). The crystallization of the AMPA-type receptor GluA2 required carboxyl-terminal and internal deletion, removal of potential glycosylation sites, alanine substitution, and specific mutation (Sobolevsky et al., 2009).

Structure Determination

Since the invention of the oscillation/rotation protocol (Arndt and Wonacott, 1977), it has quickly become the primary X-ray diffraction data collection method on single crystals. Powerful rotating anode generators have been improved continuously to generate X-rays with high intensity. Synchrotron radiation, which supplies X-rays with tunable wavelength and much higher X-ray flux than home source, was brought to the attention of biological crystallographers (Rosenbaum and Holmes, 1971) and quickly became an intimate partner of structural biology. Fast-readout large-area X-ray detectors have been critically important for data collection. Synchrotron beamlines are provided with charge-coupled detectors (CCDs), which make data collection much faster and more accurate than ever before (Phillips et al., 1993). More recently, the concept of shutter-free data collection has been made possible by the pixel array detectors (Pilatus 6M) (Kraft et al., 2009). Cryogenic protection by flash freezing crystals at ~ 100 K greatly reduced radiation damage (Rodgers, 1994), which allows complete data collection from a single crystal, instead of data merging from multiple crystals in the past.

Software for X-ray data collection and processing, structure determination, and model building has seen major improvements in the last two decades. Most notably, structure determination based on a single crystal became possible with development of the multiwavelength anomalous diffraction (MAD) method (Phillips and Hodgson, 1980; Yang et al., 1990). The ease of selenomethionine incorporation into recombinant proteins and the powerful synchrotron radiation have made MAD the preferred method for de novo protein structure determination. Ribonuclease H was the first protein whose structure was successfully determined by selenium-based MAD analysis (Yang et al., 1990). Free R factor, introduced as an objective criteria for structural cross-validation (Brünger, 1992), prevents overrefinement and has quickly become a key parameter for all X-ray structures. Free R factor was also instrumental to implementation of maximum likelihood target functions in crystallographic programs (Pannu and Read, 1996). In 1994, the CCP4 (collaborative computational project, number 4) suite was established, which collected a number of programs for various macromolecular structure determination methods by X-ray crystallography (Collaborative, 1994). Another structure determination package CNS (Brünger et al., 1998),

superseded by PHENIX (Adams et al., 2002), not only provides all necessary programs for X-ray structure solution but also incorporates the refinement method of simulated-annealing after X-plor (Brunger, 1990). Structural genomics demanded high-throughput crystallography, with a number of automated structure solution pipelines established worldwide through integration of existing programs for automation.

The use of MIR demands generation of heavy atom derivatives of the crystals. The heavy atom agents “magic seven” (Boggon and Shapiro, 2000) and “membrane’s eleven” (Morth et al., 2006) were recommended for crystals of soluble and membrane proteins, respectively. Polynuclear metal clusters were utilized for phasing of larger protein complexes (Dauter, 2005), exemplified by the ribosomal large subunit (Ban et al., 2000). Compared to heavy metals, bromide or iodine ions can be quickly soaked into protein crystals within minutes. In addition to MIR and MAD, mainstream experimental phasing methods also include single-wavelength anomalous dispersion (SAD), single isomorphous replacement (SIR), multiple isomorphous replacement with anomalous scattering (MIRAS), and single isomorphous replacement with anomalous scattering (SIRAS). Molecular replacement (MR) is used for phasing with a homologous model. Automated model-building algorithms and molecular modeling graphics such as COOT (Emsley and Cowtan, 2004), preceded by O (Jones et al., 1991), have greatly accelerated the pace of structure determination.

Prior to the mid-1990s, it wasn’t uncommon for a PhD student in an X-ray crystallography laboratory to spend 3 to 4 years to solve the phase problem, requiring understanding of Patterson maps and resolution of the Harker sections. Consequently, students then usually needed to have certain knowledge of mathematics and physics. Nowadays, a student only needs to mount crystals correctly in the X-ray beam, and the rest—data collection and structure determination—can be just a few clicks away. Intense X-rays at various synchrotron facilities have not only greatly expanded our investigative power on small and weakly diffracting crystals but have also made high-throughput crystallography a reality (Joachimik, 2009). Remote control of data collection at synchrotrons has eased the trouble of long-distance traveling and sleepless night fatigue.

Hard X-Ray Free Electron Laser

For most X-ray beamlines in existing synchrotrons, destruction of crystals by radiation damage accompanies the process of data collection. Ultrashort (femtoseconds), high-intensity X-ray pulses from free-electron lasers were predicted to provide useful structural information before the destruction of the sample by radiation damage (Neutze et al., 2000). This prediction was reached on the basis of computer simulation of radiation damage, taking into account photon energy, pulse length and intensity, and sample size. A decade later, the proof of concept was established with image construction on diffractions collected on individual virus particles (Seibert et al., 2011) and an electron density map at 8.5 Å resolution of photosystem I, which required millions of nanocrystals and more than 3,000,000 diffraction patterns (Chapman et al., 2011). Both studies were carried out at the Linac Coherent Light Source at Stanford University, which offers the world’s first hard X-ray free electron laser (XFEL) with an X-ray intensity of $\sim 10^{12}$ photons in 10 femtoseconds, or about

10–13 orders of magnitude more intense than a regular third-generation synchrotron. The first high-resolution X-ray structure by serial femtosecond crystallography (SFX) was accomplished on lysozyme (Boutet et al., 2012), followed by de novo generation of experimental phases for lysozyme (Barends et al., 2014). This technology has been successfully applied to structure determination of a precursor cathepsin B (Redecke et al., 2013), photosystem II (Kern et al., 2013), and a human serotonin receptor (Liu et al., 2013).

The breathtaking advances of XFEL-based SFX rely on development of container-free sample spraying techniques, with initial crystal hit rate of less than 0.01%. Thus, a vast number of nanocrystals were a prerequisite for such technology, although improvement is underway. By design of the SFX technology, all X-ray diffraction patterns collected are still, partial images, which demand improvements in data processing software (Hattne et al., 2014). The fact of one image per crystal also mandates the isomorphous nature of the microcrystals. These challenges, together with the exceedingly high costs associated with construction of XFEL facility, may qualify the near-future application of SFX to limited pragmatic impact. Nonetheless, as the ultimate technology for determination of macromolecular structure in aqueous solution at room temperature, XFEL is destined to dominate in the future.

Molecular Dynamics

Crystal structures represent static snapshots of macromolecules—averaged temporally and spatially over myriad of subtle conformational differences. In real life, however, these molecules may interconvert among a few distinct conformations to accomplish their biological functions. Thankfully, these distinct conformational states can be trapped by alteration of crystallization condition, inclusion of specific ligand, and/or mutation of specific amino acids. At times of difficulty in capturing one or more of these conformational states, molecular dynamics (MD) simulation came to the rescue. A super machine named Anton extended the timescale of MD simulation to millisecond and therefore enabled the observation of large-scale conformational changes of macromolecules in silico (Lindorff-Larsen et al., 2011; Shaw et al., 2010). Using this approach, the transition between different conformational states was observed for a number of proteins.

In addition to identifying putative conformational changes, MD simulation also helps reveal insights that are sometimes unnoted by structural analysis. Simulation of the c-Src and Hck kinases revealed an important role for the linker sequence between SH3 and SH2 domains, which helps maintain the kinases in an inhibited state (Young et al., 2001). 12 ns MD simulations of the aquaglyceroporin GlpF identified the spatial and temporal probability distribution of a single file of seven to nine water molecules and defined their orientation inside the channel, which mechanistically explained the impermeability of aquaporin to protons (Tajkhorshid et al., 2002). The molecular mechanism of ion selectivity, K^+ versus Na^+ , in potassium channels was rationalized by molecular simulation and theoretical computation (Bostick and Brooks, 2007; Noskov et al., 2004).

Time-Resolved X-Ray Crystallography

Conformational dynamics of macromolecules, especially those in enzyme reactions, can be captured by time-resolved X-ray

crystallography, which uses intense, polychromatic X-ray pulses to generate Laue diffraction from a single crystal upon induction of light-initiated reaction (Neutze and Moffat, 2012). A classic example is observation of photodissociation and rebinding of carbon monoxide to myoglobin and hemoglobin (Knapp et al., 2006; Srajer et al., 1996). These time-resolved findings identify myoglobin photoproducts and serve a structural basis to spectroscopic observations and MD calculations. Technical challenges specific for Laue diffraction such as overlap of energies and spots have been effectively dealt with. Maturation of time-resolved X-ray crystallography allowed its application in fragile membrane protein crystals, as evidenced by the direct observation of light-induced changes in a photosynthetic reaction center (Burgess et al., 2003). The emergence of XFELs is transforming the field of time-resolved X-ray crystallography, despite numerous unresolved challenges. In contrast to time-resolved Laue diffraction, which requires crystals, time-resolved wide-angle X-ray scattering (WAXS) allows study of macromolecules in solution and has been successfully applied to a number of proteins, including myoglobin and hemoglobin.

Concluding Remark

X-ray crystallography seemed to be at its pinnacle in the early 1990s, when I, as a motivated graduate student, felt extremely excited by the rapid emergence of crystal structures on biologically important macromolecules. The forecast was that structural biology by X-ray crystallography would become saturated within a few years, leaving structural biologists in an awkward situation of excess in supply. Another prediction, believed by many in the late 1990s, would have structural genomics—automated high-throughput X-ray crystallography—dominate the research landscape. Thankfully, neither prediction was remotely close to fact. X-ray crystallography is alive and well and continues to play an extremely important role in deciphering the secrets of life. In essence, life is all about function and structure, with the former determined by the latter. In the past five decades, structural biology through X-ray crystallography has provided important mechanistic insights into every discipline of modern experimental biology. A large proportion of all display items in the seventh edition of the Jeremy Berg and Lubert Stryer *Biochemistry* textbook are structural images, mostly derived from X-ray crystallographic analyses. X-ray crystallography is at its pinnacle. What follows pinnacle must be a downhill path—would this be true for X-ray crystallography?

In response to a question raised by a motivated, biophysics-oriented student, Doug Rees of Caltech commented at a special seminar at the end of 2013 that “If I were to choose a lab to join for my PhD thesis research, I would probably go for an EM lab.” This statement echoes the emotions of those structural biologists who have been so comfortable with X-ray crystallography as the method of choice for elucidating three-dimensional structures of macromolecules. The recent technological advances in cryo-EM, single-particle analysis have sent shockwaves through the entire structural biology community. Near-atomic resolution EM structures have been obtained not only for the megacomplexes such as ribosomes and viruses (Amunts et al., 2014; Zhang et al., 2010) but also for relatively small protein particles exemplified by the TRP channel (Liao et al., 2013). Recently, a

4.5 Å resolution density map was generated for human γ -secretase (Lu et al., 2014), which has a calculated protein molecular weight of only 160 kDa with no symmetry. In the 1990s, the Cold Spring Harbor Laboratory Course on X-ray Crystallography was extremely popular for NMR spectroscopists. That signaled a trend that is still unfolding today. In the 2014 Kuo KH Summer School of Electron Microscopy in Shanghai, a number of established X-ray crystallographers were enrolled as students. There is no doubt that a new trend is in place and will change the pace of structural biology for many years to come.

John Burdon Sanderson Haldane, one of the greatest evolutionary biologists of all times, once declared in his essay, *The Future of Biology*, that “... the future will make any detailed predictions look rather silly.” Indeed, the eternal truth in research is that the unexpected always happens. Within the past 2 years, cryo-EM has emerged as a competitive and perhaps even more favored tool for elucidation of macromolecular assemblies with molecular weight of more than 300 kDa. In the foreseeable future, advances in sample preparation and image acquisition will likely expand the advantages of cryo-EM over X-ray crystallography into complexes with smaller molecular weights. On the other hand, technological development of XFEL may strongly affect the comparison of EM versus X-ray. After all, the ultimate goal of XFEL is to make reconstruction of single-molecule scattering possible, namely to visualize single molecules with atomic details in solution. Regardless of these scenarios, X-ray crystallography will continue to dominate structure determination for many years to come, owing to its mature methodology, high resolution, convenient accessibility worldwide, and a vast number of experienced users.

ACKNOWLEDGMENTS

I apologize to those colleagues whose important contributions are not cited in this review due to space limitations. I thank Jiawei Wang, Haipeng Gong, Hongwei Wang, Haitao Li, Frank Yan, Jiawei Wu, and Xinquan Wang for comments and members of my laboratory, particularly Ruobing Ren, Qi Hu, Tian Xie, Dong Deng, Shangyu Dang, Linfeng Sun, and Peilong Lu, for discussion.

REFERENCES

- Abate, A.A., Pentimalli, F., Esposito, L., and Giordano, A. (2013). ATP-noncompetitive CDK inhibitors for cancer therapy: an overview. *Expert Opin. Investig. Drugs* 22, 895–906.
- Abrahams, J.P., Leslie, A.G., Lutter, R., and Walker, J.E. (1994). Structure at 2.8 Å resolution of F1-ATPase from bovine heart mitochondria. *Nature* 370, 621–628.
- Abramson, J., Smirnova, I., Kasho, V., Verner, G., Kaback, H.R., and Iwata, S. (2003). Structure and mechanism of the lactose permease of *Escherichia coli*. *Science* 301, 610–615.
- Acharya, R., Fry, E., Stuart, D., Fox, G., Rowlands, D., and Brown, F. (1989). The three-dimensional structure of foot-and-mouth disease virus at 2.9 Å resolution. *Nature* 337, 709–716.
- Adams, P.D., Grosse-Kunstleve, R.W., Hung, L.-W., Ioerger, T.R., McCoy, A.J., Moriarty, N.W., Read, R.J., Sacchettini, J.C., Sauter, N.K., and Terwilliger, T.C. (2002). PHENIX: building new software for automated crystallographic structure determination. *Acta Crystallogr. D Biol. Crystallogr.* 58, 1948–1954.
- Agah, S., and Faham, S. (2012). Crystallization of membrane proteins in bilayers. *Methods Mol. Biol.* 914, 3–16.

- Aggarwal, A.K., Rodgers, D.W., Drott, M., Ptashne, M., and Harrison, S.C. (1988). Recognition of a DNA operator by the repressor of phage 434: a view at high resolution. *Science* 242, 899–907.
- Alattia, J.R., Matasci, M., Dimitrov, M., Aeschbach, L., Balasubramanian, S., Hacker, D.L., Wurm, F.M., and Fraering, P.C. (2013). Highly efficient production of the Alzheimer's γ -secretase integral membrane protease complex by a multi-gene stable integration approach. *Biotechnol. Bioeng.* 110, 1995–2005.
- Allen, J.P., Feher, G., Yeates, T.O., Komiya, H., and Rees, D.C. (1987). Structure of the reaction center from *Rhodobacter sphaeroides* R-26: the protein subunits. *Proc. Natl. Acad. Sci. USA* 84, 6162–6166.
- Amunts, A., Brown, A., Bai, X.C., Ll  cer, J.L., Hussain, T., Emsley, P., Long, F., Murshudov, G., Scheres, S.H., and Ramakrishnan, V. (2014). Structure of the yeast mitochondrial large ribosomal subunit. *Science* 343, 1485–1489.
- Anderson, J.E., Ptashne, M., and Harrison, S.C. (1987). Structure of the repressor-operator complex of bacteriophage 434. *Nature* 326, 846–852.
- Arndt, U.W., and Wonacott, A. (1977). Rotation method in crystallography (North-Holland: Amsterdam).
- Assenberg, R., Wan, P.T., Geisse, S., and Mayr, L.M. (2013). Advances in recombinant protein expression for use in pharmaceutical research. *Curr. Opin. Struct. Biol.* 23, 393–402.
- Baase, W.A., Liu, L., Tronrud, D.E., and Matthews, B.W. (2010). Lessons from the lysozyme of phage T4. *Protein Sci.* 19, 631–641.
- Ban, N., Nissen, P., Hansen, J., Moore, P.B., and Steitz, T.A. (2000). The complete atomic structure of the large ribosomal subunit at 2.4 Å resolution. *Science* 289, 905–920.
- Banner, D.W., D'Arcy, A., Janes, W., Gentz, R., Schoenfeld, H.J., Broger, C., Loetscher, H., and Lesslauer, W. (1993). Crystal structure of the soluble human 55 kD TNF receptor-human TNF beta complex: implications for TNF receptor activation. *Cell* 73, 431–445.
- Barends, T.R., Foucar, L., Botha, S., Doak, R.B., Shoeman, R.L., Nass, K., Koglin, J.E., Williams, G.J., Boutet, S., Messerschmidt, M., and Schlichting, I. (2014). De novo protein crystal structure determination from X-ray free-electron laser data. *Nature* 505, 244–247.
- Bass, R.B., Strop, P., Barclay, M., and Rees, D.C. (2002). Crystal structure of *Escherichia coli* MscS, a voltage-modulated and mechanosensitive channel. *Science* 298, 1582–1587.
- Beck, F., Unverdorben, P., Bohn, S., Schweitzer, A., Pfeifer, G., Sakata, E., Nickell, S., Plitzko, J.M., Villa, E., Baumeister, W., and Förster, F. (2012). Near-atomic resolution structural model of the yeast 26S proteasome. *Proc. Natl. Acad. Sci. USA* 109, 14870–14875.
- Ben-Shem, A., Fass, D., and Bibi, E. (2007). Structural basis for intramembrane proteolysis by rhomboid serine proteases. *Proc. Natl. Acad. Sci. USA* 104, 462–466.
- Blake, C.C., Koenig, D.F., Mair, G.A., North, A.C., Phillips, D.C., and Sarma, V.R. (1965). Structure of hen egg-white lysozyme. A three-dimensional Fourier synthesis at 2 Å resolution. *Nature* 206, 757–761.
- Bloomer, A.C., Champness, J.N., Bricogne, G., Staden, R., and Klug, A. (1978). Protein disk of tobacco mosaic virus at 2.8 Å resolution showing the interactions within and between subunits. *Nature* 276, 362–368.
- Boggon, T.J., and Shapiro, L. (2000). Screening for phasing atoms in protein crystallography. *Structure* 8, R143–R149.
- Bonneau, F., Basquin, J., Ebert, J., Lorentzen, E., and Conti, E. (2009). The yeast exosome functions as a macromolecular cage to channel RNA substrates for degradation. *Cell* 139, 547–559.
- Bostick, D.L., and Brooks, C.L., 3rd. (2007). Selectivity in K⁺ channels is due to topological control of the permeant ion's coordinated state. *Proc. Natl. Acad. Sci. USA* 104, 9260–9265.
- Boutet, S., Lomb, L., Williams, G.J., Barends, T.R., Aquila, A., Doak, R.B., Weierstall, U., DePonte, D.P., Steinbrener, J., Shoeman, R.L., et al. (2012). High-resolution protein structure determination by serial femtosecond crystallography. *Science* 337, 362–364.
- Bragg, W.L. (1913). The diffraction of short electromagnetic waves by a crystal. *Proceedings of the Cambridge Philosophical Society* 17, 43–57.
- Brotherton, D.H., Dhanaraj, V., Wick, S., Brizuela, L., Dommelle, P.J., Volyanik, E., Xu, X., Parisini, E., Smith, B.O., Archer, S.J., et al. (1998). Crystal structure of the complex of the cyclin D-dependent kinase Cdk6 bound to the cell-cycle inhibitor p19INK4d. *Nature* 395, 244–250.
- Brown, M.S., Ye, J., Rawson, R.B., and Goldstein, J.L. (2000). Regulated intramembrane proteolysis: a control mechanism conserved from bacteria to humans. *Cell* 100, 391–398.
- Br  nger, A. (1990). XPLOR Manual, v 2.1 (New Haven: Yale University).
- Br  nger, A.T. (1992). Free R value: a novel statistical quantity for assessing the accuracy of crystal structures. *Nature* 355, 472–475.
- Br  nger, A.T., Adams, P.D., Clore, G.M., DeLano, W.L., Gros, P., Grosse-Kunstleve, R.W., Jiang, J.-S., Kuszewski, J., Nilges, M., Pannu, N.S., et al. (1998). Crystallography & NMR system: A new software suite for macromolecular structure determination. *Acta Crystallogr. D Biol. Crystallogr.* 54, 905–921.
- Burgess, A.W., Cho, H.S., Eigenbrot, C., Ferguson, K.M., Garrett, T.P., Leahy, D.J., Lemmon, M.A., Sliwkowski, M.X., Ward, C.W., and Yokoyama, S. (2003). An open-and-shut case? Recent insights into the activation of EGF/ErbB receptors. *Mol. Cell* 12, 541–552.
- Caffrey, M. (2009). Crystallizing membrane proteins for structure determination: use of lipidic mesophases. *Annu. Rev. Biophys.* 38, 29–51.
- Chang, G., Spencer, R.H., Lee, A.T., Barclay, M.T., and Rees, D.C. (1998). Structure of the MscL homolog from *Mycobacterium tuberculosis*: a gated mechanosensitive ion channel. *Science* 282, 2220–2226.
- Chapman, H.N., Fromme, P., Barty, A., White, T.A., Kirian, R.A., Aquila, A., Hunter, M.S., Schulz, J., DePonte, D.P., Weierstall, U., et al. (2011). Femtosecond X-ray protein nanocrystallography. *Nature* 470, 73–77.
- Chen, S., Oldham, M.L., Davidson, A.L., and Chen, J. (2013). Carbon catabolite repression of the maltose transporter revealed by X-ray crystallography. *Nature* 499, 364–368.
- Chen, L., Duerr, K.L., and Gouaux, E. (2014). X-ray structures of AMPA receptor-cone snail toxin complexes illuminate activation mechanism. *Science* 345, 1021–1026.
- Cherezov, V., Rosenbaum, D.M., Hanson, M.A., Rasmussen, S.G., Thian, F.S., Kobilka, T.S., Choi, H.J., Kuhn, P., Weis, W.I., Kobilka, B.K., and Stevens, R.C. (2007). High-resolution crystal structure of an engineered human beta2-adrenergic G protein-coupled receptor. *Science* 318, 1258–1265.
- Cherezov, V., Abola, E., and Stevens, R.C. (2010). Recent progress in the structure determination of GPCRs, a membrane protein family with high potential as pharmaceutical targets. *Methods Mol. Biol.* 654, 141–168.
- Cho, H.S., Mason, K., Ramyar, K.X., Stanley, A.M., Gabelli, S.B., Denney, D.W., Jr., and Leahy, D.J. (2003). Structure of the extracellular region of HER2 alone and in complex with the Herceptin Fab. *Nature* 421, 756–760.
- Collaborative, C.P.; Collaborative Computational Project, Number 4 (1994). The CCP4 suite: programs for protein crystallography. *Acta Crystallogr. D Biol. Crystallogr.* 50, 760–763.
- Cramer, P. (2014). A tale of chromatin and transcription in 100 structures. *Cell* 159, this issue, 985–994.
- Dang, S., Sun, L., Huang, Y., Lu, F., Liu, Y., Gong, H., Wang, J., and Yan, N. (2010). Structure of a fucose transporter in an outward-open conformation. *Nature* 467, 734–738.
- Dauter, Z. (2005). Use of polynuclear metal clusters in protein crystallography. *C. R. Chim.* 8, 1808–1814.
- Dawson, R.J., and Locher, K.P. (2006). Structure of a bacterial multidrug ABC transporter. *Nature* 443, 180–185.
- De Bondt, H.L., Rosenblatt, J., Jancarik, J., Jones, H.D., Morgan, D.O., and Kim, S.H. (1993). Crystal structure of cyclin-dependent kinase 2. *Nature* 363, 595–602.
- de Vos, A.M., Ultsch, M., and Kossiakoff, A.A. (1992). Human growth hormone and extracellular domain of its receptor: crystal structure of the complex. *Science* 255, 306–312.

- Deisenhofer, J., Epp, O., Miki, K., Huber, R., and Michel, H. (1985). Structure of the protein subunits in the photosynthetic reaction centre of *Rhodospseudomonas viridis* at 3 Å resolution. *Nature* 318, 618–624.
- Deng, D., Xu, C., Sun, P., Wu, J., Yan, C., Hu, M., and Yan, N. (2014). Crystal structure of the human glucose transporter GLUT1. *Nature* 510, 121–125.
- Doyle, D.A., Lee, A., Lewis, J., Kim, E., Sheng, M., and MacKinnon, R. (1996). Crystal structures of a complexed and peptide-free membrane protein-binding domain: molecular basis of peptide recognition by PDZ. *Cell* 85, 1067–1076.
- Doyle, D.A., Morais Cabral, J., Pfuetzner, R.A., Kuo, A., Gulbis, J.M., Cohen, S.L., Chait, B.T., and MacKinnon, R. (1998). The structure of the potassium channel: molecular basis of K⁺ conduction and selectivity. *Science* 280, 69–77.
- Drenth, J., Jansonius, J.N., Koekoek, R., Swen, H.M., and Wolthers, B.G. (1968). Structure of papain. *Nature* 218, 929–932.
- Dürr, K.L., Chen, L., Stein, R.A., De Zorzi, R., Folea, I.M., Walz, T., Mchaourab, H.S., and Gouaux, E. (2014). Structure and dynamics of AMPA receptor GluA2 in resting, pre-open, and desensitized states. *Cell* 158, 778–792.
- Emsley, P., and Cowtan, K. (2004). Coot: model-building tools for molecular graphics. *Acta Crystallogr. D Biol. Crystallogr.* 60, 2126–2132.
- Faham, S., Watanabe, A., Besserer, G.M., Cascio, D., Specht, A., Hirayama, B.A., Wright, E.M., and Abramson, J. (2008). The crystal structure of a sodium galactose transporter reveals mechanistic insights into Na⁺/sugar symport. *Science* 321, 810–814.
- Fang, Y., Jayaram, H., Shane, T., Kolmakova-Partensky, L., Wu, F., Williams, C., Xiong, Y., and Miller, C. (2009). Structure of a prokaryotic virtual proton pump at 3.2 Å resolution. *Nature* 460, 1040–1043.
- Feng, L., Yan, H., Wu, Z., Yan, N., Wang, Z., Jeffrey, P.D., and Shi, Y. (2007). Structure of a site-2 protease family intramembrane metalloprotease. *Science* 318, 1608–1612.
- Ferguson, K.M., Lemmon, M.A., Schlessinger, J., and Sigler, P.B. (1995). Structure of the high affinity complex of inositol trisphosphate with a phospholipase C pleckstrin homology domain. *Cell* 83, 1037–1046.
- Finch, J.T., Lutter, L.C., Rhodes, D., Brown, R.S., Rushton, B., Levitt, M., and Klug, A. (1977). Structure of nucleosome core particles of chromatin. *Nature* 269, 29–36.
- Fischer, E.H., and Krebs, E.G. (1955). Conversion of phosphorylase b to phosphorylase a in muscle extracts. *J. Biol. Chem.* 216, 121–132.
- Franklin, M.C., Carey, K.D., Vajdos, F.F., Leahy, D.J., de Vos, A.M., and Sliwkowski, M.X. (2004). Insights into ErbB signaling from the structure of the ErbB2-pertuzumab complex. *Cancer Cell* 5, 317–328.
- Fu, D., Libson, A., Miercke, L.J., Weitzman, C., Nollert, P., Krucinski, J., and Stroud, R.M. (2000). Structure of a glycerol-conducting channel and the basis for its selectivity. *Science* 290, 481–486.
- Gao, X., Lu, F., Zhou, L., Dang, S., Sun, L., Li, X., Wang, J., and Shi, Y. (2009). Structure and mechanism of an amino acid antiporter. *Science* 324, 1565–1568.
- Gao, X., Zhou, L., Jiao, X., Lu, F., Yan, C., Zeng, X., Wang, J., and Shi, Y. (2010). Mechanism of substrate recognition and transport by an amino acid antiporter. *Nature* 463, 828–832.
- Garboczi, D.N., Ghosh, P., Utz, U., Fan, Q.R., Biddison, W.E., and Wiley, D.C. (1996). Structure of the complex between human T-cell receptor, viral peptide and HLA-A2. *Nature* 384, 134–141.
- Garcia, K.C., Degano, M., Stanfield, R.L., Brunmark, A., Jackson, M.R., Peterson, P.A., Teyton, L., and Wilson, I.A. (1996). An alphabeta T cell receptor structure at 2.5 Å and its orientation in the TCR-MHC complex. *Science* 274, 209–219.
- Garrett, T.P., McKern, N.M., Lou, M., Elleman, T.C., Adams, T.E., Lovrecz, G.O., Zhu, H.J., Walker, F., Frenkel, M.J., Hoyne, P.A., et al. (2002). Crystal structure of a truncated epidermal growth factor receptor extracellular domain bound to transforming growth factor alpha. *Cell* 110, 763–773.
- Gerber, S., Comellas-Bigler, M., Goetz, B.A., and Locher, K.P. (2008). Structural basis of trans-inhibition in a molybdate/tungstate ABC transporter. *Science* 321, 246–250.
- Gonen, T., Cheng, Y., Sliz, P., Hiroaki, Y., Fujiyoshi, Y., Harrison, S.C., and Walz, T. (2005). Lipid-protein interactions in double-layered two-dimensional AQP0 crystals. *Nature* 438, 633–638.
- Grigorieff, N., Ceska, T.A., Downing, K.H., Baldwin, J.M., and Henderson, R. (1996). Electron-crystallographic refinement of the structure of bacteriorhodopsin. *J. Mol. Biol.* 259, 393–421.
- Groll, M., Ditzel, L., Löwe, J., Stock, D., Bochtler, M., Bartunik, H.D., and Huber, R. (1997). Structure of 20S proteasome from yeast at 2.4 Å resolution. *Nature* 386, 463–471.
- Harrison, S.C. (1969). Structure of tomato bushy stunt virus. I. The spherically averaged electron density. *J. Mol. Biol.* 42, 457–483.
- Harrison, S.C., and Jack, A. (1975). Structure of tomato bushy stunt virus. Three-dimensional x-ray diffraction analysis at 16 Å resolution. *J. Mol. Biol.* 97, 173–191.
- Harrison, S.C., Olson, A.J., Schutt, C.E., Winkler, F.K., and Bricogne, G. (1978). Tomato bushy stunt virus at 2.9 Å resolution. *Nature* 276, 368–373.
- Hattne, J., Echols, N., Tran, R., Kern, J., Gildea, R.J., Brewster, A.S., Alonso-Mori, R., Glöckner, C., Hellmich, J., Laksmono, H., et al. (2014). Accurate macromolecular structures using minimal measurements from X-ray free-electron lasers. *Nat. Methods* 11, 545–548.
- Huang, Y., Lemieux, M.J., Song, J., Auer, M., and Wang, D.N. (2003). Structure and mechanism of the glycerol-3-phosphate transporter from *Escherichia coli*. *Science* 301, 616–620.
- Hvorup, R.N., Goetz, B.A., Niederer, M., Hollenstein, K., Perozo, E., and Locher, K.P. (2007). Asymmetry in the structure of the ABC transporter-binding protein complex BtuCD-BtuF. *Science* 317, 1387–1390.
- Jardetzky, O. (1966). Simple allosteric model for membrane pumps. *Nature* 211, 969–970.
- Jeffrey, P.D., Russo, A.A., Polyak, K., Gibbs, E., Hurwitz, J., Massagué, J., and Pavletich, N.P. (1995). Mechanism of CDK activation revealed by the structure of a cyclinA-CDK2 complex. *Nature* 376, 313–320.
- Jiang, Y., Lee, A., Chen, J., Cadene, M., Chait, B.T., and MacKinnon, R. (2002). Crystal structure and mechanism of a calcium-gated potassium channel. *Nature* 417, 515–522.
- Jiang, Y., Ruta, V., Chen, J., Lee, A., and MacKinnon, R. (2003). The principle of gating charge movement in a voltage-dependent K⁺ channel. *Nature* 423, 42–48.
- Joachimik, A. (2009). High-throughput crystallography for structural genomics. *Curr. Opin. Struct. Biol.* 19, 573–584.
- Jones, T.A., Zou, J.-Y., Cowan, S.W., and Kjeldgaard, M. (1991). Improved methods for building protein models in electron density maps and the location of errors in these models. *Acta Crystallogr. A* 47, 110–119.
- Kaback, H.R. (2005). Structure and mechanism of the lactose permease. *C. R. Biol.* 328, 557–567.
- Kadaba, N.S., Kaiser, J.T., Johnson, E., Lee, A., and Rees, D.C. (2008). The high-affinity *E. coli* methionine ABC transporter: structure and allosteric regulation. *Science* 321, 250–253.
- Karakas, E., and Furukawa, H. (2014). Crystal structure of a heterotetrameric NMDA receptor ion channel. *Science* 344, 992–997.
- Kartha, G., Bello, J., and Harker, D. (1967). Tertiary structure of ribonuclease. *Nature* 213, 862–865.
- Kendrew, J.C., Bodo, G., Dintzis, H.M., Parrish, R.G., Wyckoff, H., and Phillips, D.C. (1958). A three-dimensional model of the myoglobin molecule obtained by x-ray analysis. *Nature* 181, 662–666.
- Kendrew, J.C., Dickerson, R.E., Strandberg, B.E., Hart, R.G., Davies, D.R., Phillips, D.C., and Shore, V.C. (1960). Structure of myoglobin: A three-dimensional Fourier synthesis at 2 Å resolution. *Nature* 185, 422–427.
- Kern, J., Alonso-Mori, R., Tran, R., Hattne, J., Gildea, R.J., Echols, N., Glöckner, C., Hellmich, J., Laksmono, H., Sierra, R.G., et al. (2013). Simultaneous

- femtosecond X-ray spectroscopy and diffraction of photosystem II at room temperature. *Science* 340, 491–495.
- Khare, D., Oldham, M.L., Orelle, C., Davidson, A.L., and Chen, J. (2009). Alternating access in maltose transporter mediated by rigid-body rotations. *Mol. Cell* 33, 528–536.
- Kim, C., Xuong, N.H., and Taylor, S.S. (2005). Crystal structure of a complex between the catalytic and regulatory (R α) subunits of PKA. *Science* 307, 690–696.
- Knapp, J.E., Pahl, R., Srajer, V., and Royer, W.E., Jr. (2006). Allosteric action in real time: time-resolved crystallographic studies of a cooperative dimeric hemoglobin. *Proc. Natl. Acad. Sci. USA* 103, 7649–7654.
- Knighton, D.R., Zheng, J.H., Ten Eyck, L.F., Ashford, V.A., Xuong, N.H., Taylor, S.S., and Sowadski, J.M. (1991). Crystal structure of the catalytic subunit of cyclic adenosine monophosphate-dependent protein kinase. *Science* 253, 407–414.
- Kobilka, B.K., Dixon, R.A., Frielle, T., Dohman, H.G., Bolanowski, M.A., Sigal, I.S., Yang-Feng, T.L., Francke, U., Caron, M.G., and Lefkowitz, R.J. (1987a). cDNA for the human beta 2-adrenergic receptor: a protein with multiple membrane-spanning domains and encoded by a gene whose chromosomal location is shared with that of the receptor for platelet-derived growth factor. *Proc. Natl. Acad. Sci. USA* 84, 46–50.
- Kobilka, B.K., Matsui, H., Kobilka, T.S., Yang-Feng, T.L., Francke, U., Caron, M.G., Lefkowitz, R.J., and Regan, J.W. (1987b). Cloning, sequencing, and expression of the gene coding for the human platelet alpha 2-adrenergic receptor. *Science* 238, 650–656.
- Kobilka, B.K., Kobilka, T.S., Daniel, K., Regan, J.W., Caron, M.G., and Lefkowitz, R.J. (1988). Chimeric alpha 2 β 2-adrenergic receptors: delineation of domains involved in effector coupling and ligand binding specificity. *Science* 240, 1310–1316.
- Kraft, P., Bergamaschi, A., Broennimann, Ch., Dinapoli, R., Eikenberry, E.F., Henrich, B., Johnson, I., Mozzanica, A., Schlepütz, C.M., Willmott, P.R., and Schmitt, B. (2009). Performance of single-photon-counting PILATUS detector modules. *J. Synchrotron Radiat.* 16, 368–375.
- Kühlbrandt, W., Wang, D.N., and Fujiyoshi, Y. (1994). Atomic model of plant light-harvesting complex by electron crystallography. *Nature* 367, 614–621.
- Lee, C.H., Lü, W., Michel, J.C., Goehring, A., Du, J., Song, X., and Gouaux, E. (2014). NMDA receptor structures reveal subunit arrangement and pore architecture. *Nature* 511, 191–197.
- Lemieux, M.J., Fischer, S.J., Cherney, M.M., Bateman, K.S., and James, M.N. (2007). The crystal structure of the rhomboid peptidase from *Haemophilus influenzae* provides insight into intramembrane proteolysis. *Proc. Natl. Acad. Sci. USA* 104, 750–754.
- Leung, A.K., Nagai, K., and Li, J. (2011). Structure of the spliceosomal U4 snRNP core domain and its implication for snRNP biogenesis. *Nature* 473, 536–539.
- Li, S., Schmitz, K.R., Jeffrey, P.D., Wiltzius, J.J., Kussie, P., and Ferguson, K.M. (2005). Structural basis for inhibition of the epidermal growth factor receptor by cetuximab. *Cancer Cell* 7, 301–311.
- Li, X., Dang, S., Yan, C., Gong, X., Wang, J., and Shi, Y. (2013). Structure of a presenilin family intramembrane aspartate protease. *Nature* 493, 56–61.
- Liao, M., Cao, E., Julius, D., and Cheng, Y. (2013). Structure of the TRPV1 ion channel determined by electron cryo-microscopy. *Nature* 504, 107–112.
- Lindorff-Larsen, K., Piana, S., Dror, R.O., and Shaw, D.E. (2011). How fast-folding proteins fold. *Science* 334, 517–520.
- Lingaraju, G.M., Bunker, R.D., Cavadini, S., Hess, D., Hassiepen, U., Renatus, M., Fischer, E.S., and Thomä, N.H. (2014). Crystal structure of the human COP9 signalosome. *Nature* 512, 161–165.
- Lipscomb, W.N., Hartsuck, J.A., Quirocho, F.A., and Reeke, G.N., Jr. (1969). The structure of carboxypeptidase A. IX. The x-ray diffraction results in the light of the chemical sequence. *Proc. Natl. Acad. Sci. USA* 64, 28–35.
- Liu, Q., Greimann, J.C., and Lima, C.D. (2006). Reconstitution, activities, and structure of the eukaryotic RNA exosome. *Cell* 127, 1223–1237.
- Liu, W., Wacker, D., Gati, C., Han, G.W., James, D., Wang, D., Nelson, G., Weierstall, U., Katritch, V., Barty, A., et al. (2013). Serial femtosecond crystallography of G protein-coupled receptors. *Science* 342, 1521–1524.
- Locher, K.P., Lee, A.T., and Rees, D.C. (2002). The E. coli BtuCD structure: a framework for ABC transporter architecture and mechanism. *Science* 296, 1091–1098.
- Long, S.B., Campbell, E.B., and Mackinnon, R. (2005). Crystal structure of a mammalian voltage-dependent Shaker family K⁺ channel. *Science* 309, 897–903.
- Löwe, J., Stock, D., Jap, B., Zwickl, P., Baumeister, W., and Huber, R. (1995). Crystal structure of the 20S proteasome from the archaeon *T. acidophilum* at 3.4 Å resolution. *Science* 268, 533–539.
- Lu, P., Bai, X.C., Ma, D., Xie, T., Yan, C., Sun, L., Yang, G., Zhao, Y., Zhou, R., Scheres, S.H., and Shi, Y. (2014). Three-dimensional structure of human γ -secretase. *Nature* 512, 166–170.
- Luger, K., Mäder, A.W., Richmond, R.K., Sargent, D.F., and Richmond, T.J. (1997). Crystal structure of the nucleosome core particle at 2.8 Å resolution. *Nature* 389, 251–260.
- Luo, M., Vriend, G., Kamer, G., Minor, I., Arnold, E., Rossmann, M.G., Boege, U., Scraba, D.G., Duke, G.M., and Palmenberg, A.C. (1987). The atomic structure of Mengo virus at 3.0 Å resolution. *Science* 235, 182–191.
- Lynch, T.J., Bell, D.W., Sordella, R., Gurubhagavatula, S., Okimoto, R.A., Brannigan, B.W., Harris, P.L., Haserlat, S.M., Supko, J.G., Haluska, F.G., et al. (2004). Activating mutations in the epidermal growth factor receptor underlying responsiveness of non-small-cell lung cancer to gefitinib. *N. Engl. J. Med.* 350, 2129–2139.
- Ma, D., Lu, P., Yan, C., Fan, C., Yin, P., Wang, J., and Shi, Y. (2012). Structure and mechanism of a glutamate-GABA antiporter. *Nature* 483, 632–636.
- MacKinnon, R. (2003). Potassium channels. *FEBS Lett.* 555, 62–65.
- Makino, D.L., Baumgärtner, M., and Conti, E. (2013). Crystal structure of an RNA-bound 11-subunit eukaryotic exosome complex. *Nature* 495, 70–75.
- Matthews, B.W., and Remington, S.J. (1974). The three dimensional structure of the lysozyme from bacteriophage T4. *Proc. Natl. Acad. Sci. USA* 71, 4178–4182.
- Matthews, B.W., Sigler, P.B., Henderson, R., and Blow, D.M. (1967). Three-dimensional structure of tosyl-alpha-chymotrypsin. *Nature* 214, 652–656.
- McPherson, A., and Gavira, J.A. (2014). Introduction to protein crystallization. *Acta Crystallogr. F Struct. Biol. Commun.* 70, 2–20.
- Montemayor, E.J., Curran, E.C., Liao, H.H., Andrews, K.L., Treba, C.N., Butcher, S.E., and Brow, D.A. (2014). Core structure of the U6 small nuclear ribonucleoprotein at 1.7-Å resolution. *Nat. Struct. Mol. Biol.* 21, 544–551.
- Morais-Cabral, J.H., Zhou, Y., and MacKinnon, R. (2001). Energetic optimization of ion conduction rate by the K⁺ selectivity filter. *Nature* 414, 37–42.
- Morth, J.P., Sørensen, T.L., and Nissen, P. (2006). Membrane's Eleven: heavy-atom derivatives of membrane-protein crystals. *Acta Crystallogr. D Biol. Crystallogr.* 62, 877–882.
- Musacchio, A., Noble, M., Pauptit, R., Wierenga, R., and Saraste, M. (1992). Crystal structure of a Src-homology 3 (SH3) domain. *Nature* 359, 851–855.
- Musacchio, A., Saraste, M., and Wilmanns, M. (1994). High-resolution crystal structures of tyrosine kinase SH3 domains complexed with proline-rich peptides. *Nat. Struct. Biol.* 1, 546–551.
- Nagar, B., Bornmann, W.G., Pellicena, P., Schindler, T., Veach, D.R., Miller, W.T., Clarkson, B., and Kuriyan, J. (2002). Crystal structures of the kinase domain of c-Abl in complex with the small molecule inhibitors PD173955 and imatinib (STI-571). *Cancer Res.* 62, 4236–4243.
- Neutze, R., and Moffat, K. (2012). Time-resolved structural studies at synchrotrons and X-ray free electron lasers: opportunities and challenges. *Curr. Opin. Struct. Biol.* 22, 651–659.
- Neutze, R., Wouts, R., van der Spoel, D., Weckert, E., and Hajdu, J. (2000). Potential for biomolecular imaging with femtosecond X-ray pulses. *Nature* 406, 752–757.

- Newstead, S., Drew, D., Cameron, A.D., Postis, V.L., Xia, X., Fowler, P.W., Ingram, J.C., Carpenter, E.P., Sansom, M.S., McPherson, M.J., et al. (2011). Crystal structure of a prokaryotic homologue of the mammalian oligopeptide-proton symporters, PepT1 and PepT2. *EMBO J.* 30, 417–426.
- Noskov, S.Y., Bernèche, S., and Roux, B. (2004). Control of ion selectivity in potassium channels by electrostatic and dynamic properties of carbonyl ligands. *Nature* 431, 830–834.
- Ogiso, H., Ishitani, R., Nureki, O., Fukai, S., Yamanaka, M., Kim, J.H., Saito, K., Sakamoto, A., Inoue, M., Shirouzu, M., and Yokoyama, S. (2002). Crystal structure of the complex of human epidermal growth factor and receptor extracellular domains. *Cell* 110, 775–787.
- Oldham, M.L., and Chen, J. (2011a). Crystal structure of the maltose transporter in a pretranslocation intermediate state. *Science* 332, 1202–1205.
- Oldham, M.L., and Chen, J. (2011b). Snapshots of the maltose transporter during ATP hydrolysis. *Proc. Natl. Acad. Sci. USA* 108, 15152–15156.
- Oldham, M.L., Khare, D., Quiocho, F.A., Davidson, A.L., and Chen, J. (2007). Crystal structure of a catalytic intermediate of the maltose transporter. *Nature* 450, 515–521.
- Olesen, C., Sørensen, T.L., Nielsen, R.C., Møller, J.V., and Nissen, P. (2004). Dephosphorylation of the calcium pump coupled to counterion occlusion. *Science* 306, 2251–2255.
- Olesen, C., Picard, M., Winther, A.M., Gyryp, C., Morth, J.P., Oxvig, C., Møller, J.V., and Nissen, P. (2007). The structural basis of calcium transport by the calcium pump. *Nature* 450, 1036–1042.
- Otwinowski, Z., Schevitz, R.W., Zhang, R.G., Lawson, C.L., Joachimiak, A., Marmorstein, R.Q., Luisi, B.F., and Sigler, P.B. (1988). Crystal structure of trp repressor/operator complex at atomic resolution. *Nature* 335, 321–329.
- Paez, J.G., Jänne, P.A., Lee, J.C., Tracy, S., Greulich, H., Gabriel, S., Herman, P., Kaye, F.J., Lindeman, N., Boggon, T.J., et al. (2004). EGFR mutations in lung cancer: correlation with clinical response to gefitinib therapy. *Science* 304, 1497–1500.
- Palczewski, K., Kumasaka, T., Hori, T., Behnke, C.A., Motoshima, H., Fox, B.A., Le Trong, I., Teller, D.C., Okada, T., Stenkamp, R.E., et al. (2000). Crystal structure of rhodopsin: A G protein-coupled receptor. *Science* 289, 739–745.
- Pannu, N.S., and Read, R.J. (1996). Improved structure refinement through maximum likelihood. *Acta Crystallogr. A* 52, 659–668.
- Pauling, L., and Corey, R.B. (1951a). Atomic coordinates and structure factors for two helical configurations of polypeptide chains. *Proc. Natl. Acad. Sci. USA* 37, 235–240.
- Pauling, L., and Corey, R.B. (1951b). Configuration of polypeptide chains. *Nature* 168, 550–551.
- Pauling, L., and Corey, R.B. (1951c). The polypeptide-chain configuration in hemoglobin and other globular proteins. *Proc. Natl. Acad. Sci. USA* 37, 282–285.
- Pauling, L., Corey, R.B., and Branson, H.R. (1951). The structure of proteins; two hydrogen-bonded helical configurations of the polypeptide chain. *Proc. Natl. Acad. Sci. USA* 37, 205–211.
- Pavletich, N.P., Chambers, K.A., and Pabo, C.O. (1993). The DNA-binding domain of p53 contains the four conserved regions and the major mutation hot spots. *Genes Dev.* 7(12B), 2556–2564.
- Pebay-Peyroula, E., Rummel, G., Rosenbusch, J.P., and Landau, E.M. (1997). X-ray structure of bacteriorhodopsin at 2.5 angstroms from microcrystals grown in lipidic cubic phases. *Science* 277, 1676–1681.
- Perutz, M.F., Rossmann, M.G., Cullis, A.F., Muirhead, H., Will, G., and North, A.C. (1960). Structure of haemoglobin: a three-dimensional Fourier synthesis at 5.5-Å resolution, obtained by X-ray analysis. *Nature* 185, 416–422.
- Perutz, M.F., Muirhead, H., Cox, J.M., Goaman, L.C., Mathews, F.S., McGandy, E.L., and Webb, L.E. (1968a). Three-dimensional Fourier synthesis of horse oxyhaemoglobin at 2.8 Å resolution: (1) x-ray analysis. *Nature* 219, 29–32.
- Perutz, M.F., Muirhead, H., Cox, J.M., and Goaman, L.C. (1968b). Three-dimensional Fourier synthesis of horse oxyhaemoglobin at 2.8 Å resolution: the atomic model. *Nature* 219, 131–139.
- Phillips, J.C., and Hodgson, K.O. (1980). The use of anomalous scattering effects to phase diffraction patterns from macromolecules. *Acta Crystallogr. A* 36, 856–864.
- Phillips, W.C., Li, Y., Stanton, M., Xie, J., O'Mara, D., and Kalata, K. (1993). A CCD-based area detector for X-ray crystallography using synchrotron and laboratory sources. *Nucl. Instrum. Methods Phys. Res. A* 334, 621–630.
- Pinkett, H.W., Lee, A.T., Lum, P., Locher, K.P., and Rees, D.C. (2007). An inward-facing conformation of a putative metal-chelate-type ABC transporter. *Science* 315, 373–377.
- Pomeranz Krummel, D.A., Oubridge, C., Leung, A.K., Li, J., and Nagai, K. (2009). Crystal structure of human spliceosomal U1 snRNP at 5.5 Å resolution. *Nature* 458, 475–480.
- Rasmussen, S.G., Choi, H.J., Rosenbaum, D.M., Kobilka, T.S., Thian, F.S., Edwards, P.C., Burghammer, M., Ratnala, V.R., Sanishvili, R., Fischetti, R.F., et al. (2007). Crystal structure of the human beta2 adrenergic G-protein-coupled receptor. *Nature* 450, 383–387.
- Rasmussen, S.G., DeVree, B.T., Zou, Y., Kruse, A.C., Chung, K.Y., Kobilka, T.S., Thian, F.S., Chae, P.S., Pardon, E., Calinski, D., et al. (2011). Crystal structure of the β_2 adrenergic receptor-Gs protein complex. *Nature* 477, 549–555.
- Recondo, G., Diaz Canton, E., de la Vega, M., Greco, M., Recondo, G., and Valsecchi, M.E. (2014). Therapeutic options for HER-2 positive breast cancer: Perspectives and future directions. *World J. Clin. Oncol.* 5, 440–454.
- Redecke, L., Nass, K., DePonte, D.P., White, T.A., Rehders, D., Barty, A., Stellato, F., Liang, M., Barends, T.R., Boutet, S., et al. (2013). Natively inhibited Trypanosoma brucei cathepsin B structure determined by using an X-ray laser. *Science* 339, 227–230.
- Ressl, S., Terwisscha van Scheltinga, A.C., Vonrhein, C., Ott, V., and Ziegler, C. (2009). Molecular basis of transport and regulation in the Na(+)/betaine symporter BetP. *Nature* 458, 47–52.
- Richmond, T.J., Finch, J.T., Rushton, B., Rhodes, D., and Klug, A. (1984). Structure of the nucleosome core particle at 7 Å resolution. *Nature* 311, 532–537.
- Rodgers, D.W. (1994). Cryocrystallography. *Structure* 2, 1135–1140.
- Rosenbaum, G., and Holmes, K. (1971). Synchrotron radiation as a source for X-ray diffraction. *Nature* 230, 434–437.
- Rosenbaum, D.M., Cherezov, V., Hanson, M.A., Rasmussen, S.G., Thian, F.S., Kobilka, T.S., Choi, H.J., Yao, X.J., Weis, W.I., Stevens, R.C., and Kobilka, B.K. (2007). GPCR engineering yields high-resolution structural insights into beta2-adrenergic receptor function. *Science* 318, 1266–1273.
- Rossmann, M.G., Arnold, E., Erickson, J.W., Frankenberger, E.A., Griffith, J.P., Hecht, H.J., Johnson, J.E., Kamer, G., Luo, M., Mosser, A.G., et al. (1985). Structure of a human common cold virus and functional relationship to other picornaviruses. *Nature* 317, 145–153.
- Roux, B., and MacKinnon, R. (1999). The cavity and pore helices in the KcsA K⁺ channel: electrostatic stabilization of monovalent cations. *Science* 285, 100–102.
- Russo, A.A., Jeffrey, P.D., Patten, A.K., Massagué, J., and Pavletich, N.P. (1996a). Crystal structure of the p27Kip1 cyclin-dependent-kinase inhibitor bound to the cyclin A-Cdk2 complex. *Nature* 382, 325–331.
- Russo, A.A., Jeffrey, P.D., and Pavletich, N.P. (1996b). Structural basis of cyclin-dependent kinase activation by phosphorylation. *Nat. Struct. Biol.* 3, 696–700.
- Russo, A.A., Tong, L., Lee, J.O., Jeffrey, P.D., and Pavletich, N.P. (1998). Structural basis for inhibition of the cyclin-dependent kinase Cdk6 by the tumour suppressor p16INK4a. *Nature* 395, 237–243.
- Saier, M.H., Jr., Yen, M.R., Noto, K., Tamang, D.G., and Elkan, C. (2009). The Transporter Classification Database: recent advances. *Nucleic Acids Res.* 37, D274–D278.

- Savage, D.F., Egea, P.F., Robles-Colmenares, Y., O'Connell, J.D., 3rd, and Stroud, R.M. (2003). Architecture and selectivity in aquaporins: 2.5 Å X-ray structure of aquaporin Z. *PLoS Biol.* 1, E72.
- Schindler, T., Bornmann, W., Pellicena, P., Miller, W.T., Clarkson, B., and Kuriyan, J. (2000). Structural mechanism for STI-571 inhibition of abelson tyrosine kinase. *Science* 289, 1938–1942.
- Seemüller, E., Lupas, A., Stock, D., Löwe, J., Huber, R., and Baumeister, W. (1995). Proteasome from *Thermoplasma acidophilum*: a threonine protease. *Science* 268, 579–582.
- Seibert, M.M., Ekeberg, T., Maia, F.R., Svenda, M., Andreasson, J., Jönsson, O., Odić, D., Iwan, B., Rocker, A., Westphal, D., et al. (2011). Single mimivirus particles intercepted and imaged with an X-ray laser. *Nature* 470, 78–81.
- Shah, N.P., and Sawyers, C.L. (2003). Mechanisms of resistance to STI571 in Philadelphia chromosome-associated leukemias. *Oncogene* 22, 7389–7395.
- Shah, K., Liu, Y., Deirmengian, C., and Shokat, K.M. (1997). Engineering unnatural nucleotide specificity for Rous sarcoma virus tyrosine kinase to uniquely label its direct substrates. *Proc. Natl. Acad. Sci. USA* 94, 3565–3570.
- Shaw, D.E., Maragakis, P., Lindorff-Larsen, K., Piana, S., Dror, R.O., Eastwood, M.P., Bank, J.A., Jumper, J.M., Salmon, J.K., Shan, Y., and Wriggers, W. (2010). Atomic-level characterization of the structural dynamics of proteins. *Science* 330, 341–346.
- Sicheri, F., Moarefi, I., and Kuriyan, J. (1997). Crystal structure of the Src family tyrosine kinase Hck. *Nature* 385, 602–609.
- Sobolevsky, A.I., Rosconi, M.P., and Gouaux, E. (2009). X-ray structure, symmetry and mechanism of an AMPA-subtype glutamate receptor. *Nature* 462, 745–756.
- Solcan, N., Kwok, J., Fowler, P.W., Cameron, A.D., Drew, D., Iwata, S., and Newstead, S. (2012). Alternating access mechanism in the POT family of oligopeptide transporters. *EMBO J.* 31, 3411–3421.
- Sørensen, T.L., Møller, J.V., and Nissen, P. (2004). Phosphoryl transfer and calcium ion occlusion in the calcium pump. *Science* 304, 1672–1675.
- Srajer, V., Teng, T., Ursby, T., Pradervand, C., Ren, Z., Adachi, S., Schildkamp, W., Bourgeois, D., Wulff, M., and Moffat, K. (1996). Photolysis of the carbon monoxide complex of myoglobin: nanosecond time-resolved crystallography. *Science* 274, 1726–1729.
- Su, Y., Dostmann, W.R., Herberg, F.W., Durick, K., Xuong, N.H., Ten Eyck, L., Taylor, S.S., and Varughese, K.I. (1995). Regulatory subunit of protein kinase A: structure of deletion mutant with cAMP binding domains. *Science* 269, 807–813.
- Subramaniam, S., Hirai, T., and Henderson, R. (2002). From structure to mechanism: electron crystallographic studies of bacteriorhodopsin. *Philos. Trans. A, Math. Phys. Eng. Sci.* 360, 859–874.
- Sui, H., Han, B.G., Lee, J.K., Walian, P., and Jap, B.K. (2001). Structural basis of water-specific transport through the AQP1 water channel. *Nature* 414, 872–878.
- Sun, L., Zeng, X., Yan, C., Sun, X., Gong, X., Rao, Y., and Yan, N. (2012). Crystal structure of a bacterial homologue of glucose transporters GLUT1-4. *Nature* 490, 361–366.
- Sutherland, E.W., Jr., and Wosilait, W.D. (1955). Inactivation and activation of liver phosphorylase. *Nature* 175, 169–170.
- Tajkhorshid, E., Nollert, P., Jensen, M.O., Miercke, L.J., O'Connell, J., Stroud, R.M., and Schulten, K. (2002). Control of the selectivity of the aquaporin water channel family by global orientational tuning. *Science* 296, 525–530.
- Toyoshima, C., and Mizutani, T. (2004). Crystal structure of the calcium pump with a bound ATP analogue. *Nature* 430, 529–535.
- Toyoshima, C., and Nomura, H. (2002). Structural changes in the calcium pump accompanying the dissociation of calcium. *Nature* 418, 605–611.
- Toyoshima, C., Nakasako, M., Nomura, H., and Ogawa, H. (2000). Crystal structure of the calcium pump of sarcoplasmic reticulum at 2.6 Å resolution. *Nature* 405, 647–655.
- Toyoshima, C., Nomura, H., and Tsuda, T. (2004). Lumenal gating mechanism revealed in calcium pump crystal structures with phosphate analogues. *Nature* 432, 361–368.
- Waight, A.B., Love, J., and Wang, D.N. (2010). Structure and mechanism of a pentameric formate channel. *Nat. Struct. Mol. Biol.* 17, 31–37.
- Waksman, G., Kominos, D., Robertson, S.C., Pant, N., Baltimore, D., Birge, R.B., Cowburn, D., Hanafusa, H., Mayer, B.J., Overduin, M., et al. (1992). Crystal structure of the phosphotyrosine recognition domain SH2 of v-src complexed with tyrosine-phosphorylated peptides. *Nature* 358, 646–653.
- Walter, M.R., Windsor, W.T., Nagabhushan, T.L., Lundell, D.J., Lunn, C.A., Zauodny, P.J., and Narula, S.K. (1995). Crystal structure of a complex between interferon-gamma and its soluble high-affinity receptor. *Nature* 376, 230–235.
- Wang, Y., Zhang, Y., and Ha, Y. (2006). Crystal structure of a rhomboid family intramembrane protease. *Nature* 444, 179–180.
- Wang, Y., Huang, Y., Wang, J., Cheng, C., Huang, W., Lu, P., Xu, Y.N., Wang, P., Yan, N., and Shi, Y. (2009). Structure of the formate transporter FocA reveals a pentameric aquaporin-like channel. *Nature* 462, 467–472.
- Wang, T., Fu, G., Pan, X., Wu, J., Gong, X., Wang, J., and Shi, Y. (2013). Structure of a bacterial energy-coupling factor transporter. *Nature* 497, 272–276.
- Ward, A., Reyes, C.L., Yu, J., Roth, C.B., and Chang, G. (2007). Flexibility in the ABC transporter MsbA: Alternating access with a twist. *Proc. Natl. Acad. Sci. USA* 104, 19005–19010.
- Wasmuth, E.V., Janusz, K., and Lima, C.D. (2014). Structure of an Rps6-RNA exosome complex bound to poly(A) RNA. *Nature* 511, 435–439.
- Watson, J.D., and Crick, F.H. (1953). Molecular structure of nucleic acids; a structure for deoxyribose nucleic acid. *Nature* 171, 737–738.
- Weyand, S., Shimamura, T., Yajima, S., Suzuki, S., Mirza, O., Krusong, K., Carpenter, E.P., Rutherford, N.G., Hadden, J.M., O'Reilly, J., et al. (2008). Structure and molecular mechanism of a nucleobase-cation-symport-1 family transporter. *Science* 322, 709–713.
- Widdas, W.F. (1952). Inability of diffusion to account for placental glucose transfer in the sheep and consideration of the kinetics of a possible carrier transfer. *J. Physiol.* 118, 23–39.
- Wiesmann, C., Fuh, G., Christinger, H.W., Eigenbrot, C., Wells, J.A., and de Vos, A.M. (1997). Crystal structure at 1.7 Å resolution of VEGF in complex with domain 2 of the Flt-1 receptor. *Cell* 91, 695–704.
- Wing, R., Drew, H., Takano, T., Broka, C., Tanaka, S., Itakura, K., and Dickerson, R.E. (1980). Crystal structure analysis of a complete turn of B-DNA. *Nature* 287, 755–758.
- Wolberger, C., Dong, Y.C., Ptashne, M., and Harrison, S.C. (1988). Structure of a phage 434 Cro/DNA complex. *Nature* 335, 789–795.
- Wright, C.S., Alden, R.A., and Kraut, J. (1969). Structure of subtilisin BPN' at 2.5 angstrom resolution. *Nature* 221, 235–242.
- Wu, Z., Yan, N., Feng, L., Oberstein, A., Yan, H., Baker, R.P., Gu, L., Jeffrey, P.D., Urban, S., and Shi, Y. (2006). Structural analysis of a rhomboid family intramembrane protease reveals a gating mechanism for substrate entry. *Nat. Struct. Mol. Biol.* 13, 1084–1091.
- Wu, J., Brown, S.H., von Daake, S., and Taylor, S.S. (2007). PKA type IIα holoenzyme reveals a combinatorial strategy for isoform diversity. *Science* 318, 274–279.
- Wyckoff, H.W., Hardman, K.D., Allewell, N.M., Inagami, T., Johnson, L.N., and Richards, F.M. (1967). The structure of ribonuclease-S at 3.5 Å resolution. *J. Biol. Chem.* 242, 3984–3988.
- Xie, T., Yan, C., Zhou, R., Zhao, Y., Sun, L., Yang, G., Lu, P., Ma, D., and Shi, Y. (2014). Crystal structure of the γ-secretase component nicastrin. *Proc. Natl. Acad. Sci. USA* 111, 13349–13354.
- Xu, W., Harrison, S.C., and Eck, M.J. (1997a). Three-dimensional structure of the tyrosine kinase c-Src. *Nature* 385, 595–602.
- Xu, Z., Horwich, A.L., and Sigler, P.B. (1997b). The crystal structure of the asymmetric GroEL-GroES-(ADP)7 chaperonin complex. *Nature* 388, 741–750.

- Xu, K., Zhang, M., Zhao, Q., Yu, F., Guo, H., Wang, C., He, F., Ding, J., and Zhang, P. (2013). Crystal structure of a folate energy-coupling factor transporter from *Lactobacillus brevis*. *Nature* 497, 268–271.
- Yamashita, A., Singh, S.K., Kawate, T., Jin, Y., and Gouaux, E. (2005). Crystal structure of a bacterial homologue of Na⁺/Cl[−]-dependent neurotransmitter transporters. *Nature* 437, 215–223.
- Yang, W., Hendrickson, W.A., Crouch, R.J., and Satow, Y. (1990). Structure of ribonuclease H phased at 2 Å resolution by MAD analysis of the selenomethionyl protein. *Science* 249, 1398–1405.
- Yin, Y., He, X., Szewczyk, P., Nguyen, T., and Chang, G. (2006). Structure of the multidrug transporter EmrD from *Escherichia coli*. *Science* 312, 741–744.
- Young, M.A., Gonfloni, S., Superti-Furga, G., Roux, B., and Kuriyan, J. (2001). Dynamic coupling between the SH2 and SH3 domains of c-Src and Hck underlies their inactivation by C-terminal tyrosine phosphorylation. *Cell* 105, 115–126.
- Zhang, X., Gureasko, J., Shen, K., Cole, P.A., and Kuriyan, J. (2006). An allosteric mechanism for activation of the kinase domain of epidermal growth factor receptor. *Cell* 125, 1137–1149.
- Zhang, X., Jin, L., Fang, Q., Hui, W.H., and Zhou, Z.H. (2010). 3.3 Å cryo-EM structure of a nonenveloped virus reveals a priming mechanism for cell entry. *Cell* 141, 472–482.
- Zhang, P., Smith-Nguyen, E.V., Keshwani, M.M., Deal, M.S., Kornev, A.P., and Taylor, S.S. (2012). Structure and allostery of the PKA RIIB tetrameric holoenzyme. *Science* 335, 712–716.
- Zhou, Y., Morais-Cabral, J.H., Kaufman, A., and MacKinnon, R. (2001). Chemistry of ion coordination and hydration revealed by a K⁺ channel-Fab complex at 2.0 Å resolution. *Nature* 414, 43–48.
- Zhou, L., Hang, J., Zhou, Y., Wan, R., Lu, G., Yin, P., Yan, C., and Shi, Y. (2014). Crystal structures of the LSM complex bound to the 3′ end sequence of U6 small nuclear RNA. *Nature* 506, 116–120.

Genetic Variation in Human DNA Replication Timing

Amnon Koren,^{1,2} Robert E. Handsaker,^{1,2,3} Nolan Kamitaki,^{1,2,3} Rosa Karlić,⁴ Sulagna Ghosh,^{1,2,3,5} Paz Polak,^{6,7} Kevin Eggan,^{3,5} and Steven A. McCarroll^{1,2,3,*}

¹Department of Genetics, Harvard Medical School, Boston, MA 02115, USA

²Program in Medical and Population Genetics, Broad Institute of MIT and Harvard, Cambridge, MA 02142, USA

³Stanley Center for Psychiatric Research, Broad Institute of MIT and Harvard, Cambridge, MA 02142, USA

⁴Bioinformatics Group, Division of Biology, Faculty of Science, Zagreb University, 10000 Zagreb, Croatia

⁵Department of Stem Cell and Regenerative Biology, The Howard Hughes Medical Institute, Harvard Stem Cell Institute, Harvard University, Cambridge, MA 02138, USA

⁶Cancer Program, Broad Institute of MIT and Harvard, Cambridge, MA 02142, USA

⁷Department of Medicine, Cancer Center and Department of Pathology, Massachusetts General Hospital, Harvard Medical School, Boston, MA 02114, USA

*Correspondence: mccarroll@genetics.med.harvard.edu

<http://dx.doi.org/10.1016/j.cell.2014.10.025>

SUMMARY

Genomic DNA replicates in a choreographed temporal order that impacts the distribution of mutations along the genome. We show here that DNA replication timing is shaped by genetic polymorphisms that act in *cis* upon megabase-scale DNA segments. In genome sequences from proliferating cells, read depth along chromosomes reflected DNA replication activity in those cells. We used this relationship to analyze variation in replication timing among 161 individuals sequenced by the 1000 Genomes Project. Genome-wide association of replication timing with genetic variation identified 16 loci at which inherited alleles associate with replication timing. We call these “replication timing quantitative trait loci” (rtQTLs). rtQTLs involved the differential use of replication origins, exhibited allele-specific effects on replication timing, and associated with gene expression variation at megabase scales. Our results show replication timing to be shaped by genetic polymorphism and identify a means by which inherited polymorphism regulates the mutability of nearby sequences.

INTRODUCTION

Replication of eukaryotic genomes follows a strict temporal program with each chromosome containing segments of characteristic early and late replication. This program is mediated by the locations and activation timing of replication origins along each chromosome (Rhind and Gilbert, 2013). Expressed genes tend to reside in early-replicating region of the genome (Rhind and Gilbert, 2013). Compared to early phases of replication, late phases of replication are faster, less structured (Koren and McCarroll, 2014), and more mutation-prone; late-replicating loci have elevated mutation rates in the human germline (Stamatoyannopoulos et al., 2009), in somatic cells (Koren et al., 2012),

and in cancer cells (Lawrence et al., 2013). Structural mutations and chromosome fragility are also more common in late-replicating genomic regions (Koren et al., 2012; Letessier et al., 2011). At the other extreme, chromosome fragility (and consequent mutations) are also increased at specific “early replicating fragile sites” (ERFSs), a subset of early replication origins at which interference between replication and transcription leads to double strand breaks (Barlow et al., 2013; Pedersen and De, 2013; Drier et al., 2013). These aspects of genome replication are conserved all the way to prokaryotes, in which genes close to the replication origin have increased expression relative to genes close to the terminus (Slager et al., 2014; Rocha, 2008), essential genes tend to be co-oriented with the direction of replication fork progression (Rocha, 2008), and the rate of mutation gradually increases with distance from the origin (Sharp et al., 1989), although close proximity to the origin can lead to structural alterations under conditions of replication stress (Slager et al., 2014).

A genome’s elaborate program of DNA replication is therefore strongly connected to genome function and evolution and could, in principle, be an object of variation and selection itself. However, it is not known whether DNA replication timing varies among members of the same species, nor whether such variation is under genetic control. Previous studies have concluded that replication timing is globally similar among individuals of the same species (Ryba et al., 2010, 2012; Hiratani et al., 2008; Pope et al., 2011; Mukhopadhyay et al., 2014). We hypothesized that this global similarity could still in principle coexist with inter-individual variation at many individual loci and that such variation might be used to find genetic influences on replication timing.

RESULTS

DNA Replication Timing Varies among Humans

DNA replication results in dynamic changes in the copy number of each genomic locus; the earlier a locus replicates, the greater its average copy number in replicating (S phase) cells. To profile these differences across the genome, we have previously isolated G1 and S phase cells using FACS (Figure 1A), sequenced

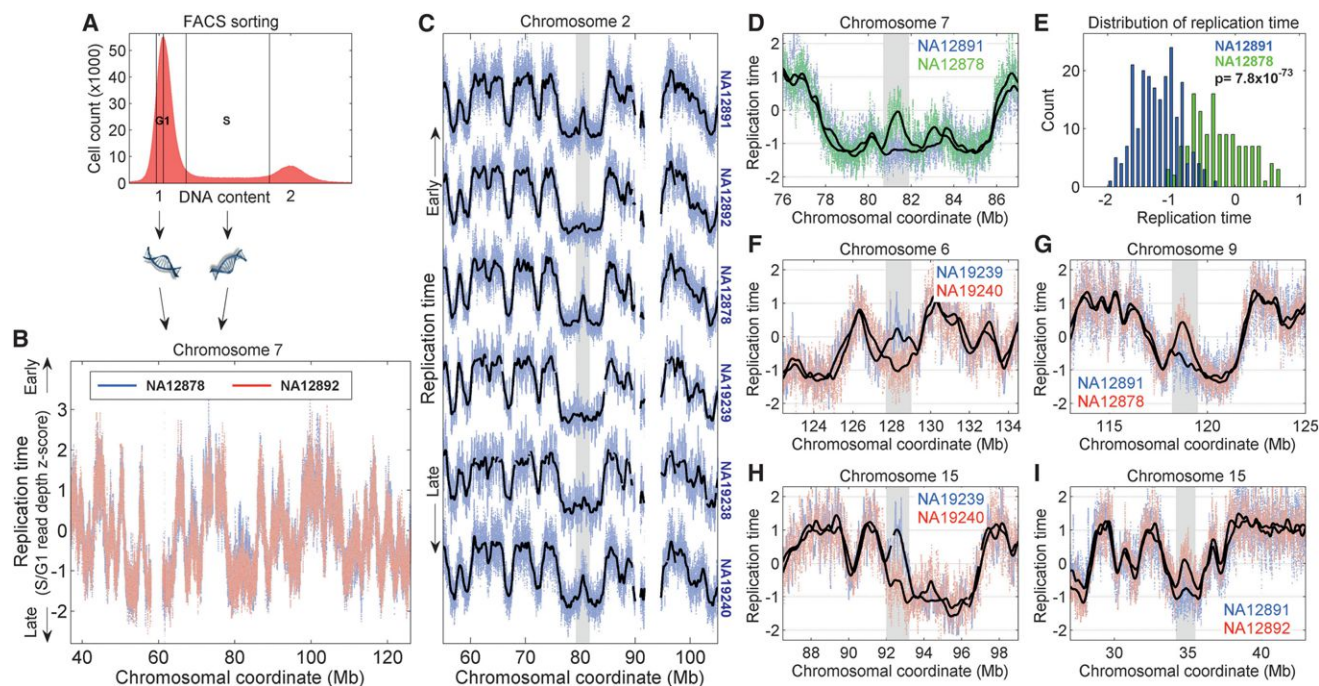


Figure 1. DNA Replication Timing Varies among Individuals at Specific Loci

(A) FACS-sorting cells by DNA content enables analysis of DNA copy number (by whole-genome sequencing) in G1 and S phase cells (adapted from Koren et al., 2012).

(B) Analysis of the ratio of DNA copy number between S and G1 phase cells along each chromosome allows the construction of replication timing profiles; early-replicating loci have a higher average copy number in S phase cells relative to late-replicating loci. Cells from different individuals show consistent replication timing programs across most of their genomes. In this and all subsequent figures, replication timing (and read depth) data are normalized to have a genome-wide mean of zero and SD of one; the y scale thus represents Z score units.

(C) A genomic locus (gray shading) exhibits interindividual variation in DNA replication timing, with only three of the six individuals exhibiting a replication origin peak structure at this locus. Black lines: smoothed replication profiles.

(D) An overlay of replication profiles from two individuals reveals a locus with variation in origin activity.

(E) The local distribution of replication timing measurements across many adjacent data windows allows statistical detection of replication timing variants. The example depicts the distributions in the genomic region shown in (D).

(F and G) Replication variants in which a replication origin (or origin cluster) is active in some individuals but inactive in others, as inferred from the presence or absence of a peak in the replication profiles.

(H and I) Replication variants in which the average utilization or activation time of a replication origin varies among individuals, as inferred from differences in peak height.

the DNA from both cell-cycle phases, and inferred replication timing from the long-range fluctuations in relative sequence abundance (the ratio of sequencing read depths from S- and G1-phase cells) along each chromosome (Figure 1B) (Koren et al., 2012). To facilitate interpretation and comparison of replication profiles, we normalize replication timing to units of SDs (Z score units, with a genome-wide average of zero and SD of one). Replication profiles provide information regarding the time of replication of each locus in the genome. They also provide the estimated locations of replication origins, which are inferred from peaks along the replication profiles, where replication is earlier than the replication of flanking sequences (Raghuraman et al., 2001; Hawkins et al., 2013); in mammalian genomes, replication peaks correspond to either single origins or clusters of closely-spaced replication origins. In previous analyses of replication timing in lymphoblastoid cell lines from six individuals, we compared the individual-averaged profiles to patterns of mutations and variation in the human genome (Koren

et al., 2012) and compared the replication profiles of female active and inactive X chromosomes (Koren and McCarroll, 2014).

We sought in the current work to better ascertain and understand interindividual variation in DNA replication. The replication profiles of the six individuals closely matched one another across most of the genome (correlation coefficients $r = 0.91$ – 0.97 among all comparisons; Figures 1B and 1C), consistent with earlier findings that profiles from different individuals are broadly similar at genomic scales (Ryba et al., 2010, 2012; Hiratani et al., 2008; Pope et al., 2011; Mukhopadhyay et al., 2014). However, at scales of several hundred kilobases, we found that specific genomic loci exhibited clear differences in replication timing among the six individuals (Figure 1C). A systematic search for replication variation identified 221 replication-variant loci (Figures 1D–1I; Extended Experimental Procedures available online), each of which spanned 0.2–1.4 Mb (mean = 0.43 Mb). At most variant loci, individuals differed qualitatively in the usage of an origin (or origin cluster) (Figures 1C, 1D, 1F, and 1G), as

inferred from the presence of a peak in the replication profile; or quantitatively in the average utilization or activation time of a replication origin (Figures 1H and 1I), as inferred from variations in the height of a peak.

This analysis indicated substantial replication timing variation among humans but did not establish whether any aspect of this variation is under genetic control. Importantly, other factors could in principle influence the observed inter-individual variations in replication timing, including epigenetic influences or even the growth states of the cells at the time of DNA harvesting and the transformation of the cells with EBV. To identify those replication phenotypes that consistently associate with specific alleles, genetic mapping requires analysis of DNA replication timing in far more individuals. However, studies of replication timing to date have involved small numbers of samples.

DNA Replication Activity Is Visible in Whole-Genome Sequence Data

Whole genome sequencing is increasingly used to study DNA sequence variation in large numbers of humans; some studies, such as the 1000 Genomes Project, use DNA samples extracted from cultured, proliferating cells (1000 Genome Project Consortium, 2012). We hypothesized that active DNA replication might be visible in such data: the presence of S phase cells in such cultures could in principle cause long-range fluctuations in DNA copy number (measured by read depth) along each chromosome, with early replicating loci contributing more DNA than late-replicating loci.

Array- and sequencing-based profiles of DNA copy number have long been known to contain megabase-scale “wave” patterns of copy number fluctuations that correlate with large-scale patterns of GC content along mammalian chromosomes (Marioni et al., 2007; Diskin et al., 2008; van de Wiel et al., 2009; Lepître et al., 2010; van Heesch et al., 2013; Aird et al., 2011). The sources of these GC-wave effects have been assumed to be technical. However, although GC content influences the efficiency of PCR amplification, GC-wave effects are present at megabase rather than subkilobase (amplicon) scales. Notably, GC content and DNA replication timing are highly correlated at megabase scales (Rhind and Gilbert, 2013), and DNA copy number is typically measured in cell populations that are derived from asynchronous, proliferating cell cultures that contain many cells in S phase. In fact, a recent study (contemporaneous with this work) noted a visual resemblance and statistical correlation between a copy number profile (derived by array CGH) and DNA replication timing profiles (Manukjan et al., 2013).

We designed a series of tests of the hypothesis that variation in sequencing coverage along chromosomes arises from true heterogeneity in DNA copy number due to ongoing DNA replication. We first tested this hypothesis using whole-genome sequence data from Phase I of the 1000 Genomes Project (1000 Genome Project Consortium, 2012), which sequenced DNA from non-synchronized, proliferating LCL cultures. For each genome analyzed, we calculated read depth along each chromosome in sliding windows of 10 kb of uniquely alignable sequence, normalized for local GC content at amplicon (400 bp) scale (Handsaker et al., 2011) (Extended Experimental Procedures). Strikingly, in most LCL-derived genome sequences, fluctuations

in read depth along each chromosome matched the LCL replication timing profiles that we had obtained by directly comparing G1 to S phase cells (Figures 2A and 2B), suggesting that they reflect true differences in underlying DNA copy number arising from replicating cells. The presence of 5%–20% of S phase cells within a cell culture was sufficient in order to yield significant signals of DNA replication timing (Figure S1E; Extended Experimental Procedures).

To further test the hypothesis that active DNA replication causes long-range fluctuations in read depth, we utilized the fact that a subset of the 1000 Genomes Project samples were derived from blood instead of LCLs. Because circulating blood cells have generally exited the cell cycle, these samples do not contain cells in S phase and should not exhibit signatures of DNA replication timing. Consistent with this hypothesis, read depth in blood-derived DNA samples lacked the strong autocorrelative patterns along chromosomes that we observed in LCL-derived DNA and was uncorrelated with profiles of replication timing (Figures 2A, 2B, and S1; Extended Experimental Procedures). In fact, we could distinguish blood-derived from LCL-derived DNA samples with 100% sensitivity and specificity, based solely on the relationship of their read depth profiles to our independent analyses of LCL replication timing (Figures 2A and S1; Table S1).

Importantly, correlations between read depth and replication timing remained strong after controlling for GC content, whereas correlations between read depth and GC content (at scales >10 kb) were negligible after controlling for replication timing (Figures 2B and S1), suggesting that previously observed correlations of read depth with GC content (at 100 kb scales) are due to DNA replication timing. Furthermore, in a principal component analysis (Patterson et al., 2006) of read depth along each of the LCL genomes, the strongest principal component, explaining 40% of the variance, corresponded to our estimate of the S phase replication content of each sample, and the chromosomal loadings of this component followed the replication timing profile (Figures 2C and 2D; Extended Experimental Procedures).

The X chromosome provided an additional strong prediction of the hypothesis that read depth at a locus reflects the replication timing of that locus. We recently found that the inactive X chromosome in females undergoes a spatially unstructured, “random” form of replication (Koren and McCarroll, 2014). In light of that finding, the hypothesis that long-range fluctuations in read depth reflect active DNA replication predicts that inter-individual correlations in read depth along each chromosome would be weaker on the X chromosome in comparisons involving female genomes. The 1000 Genomes data abundantly confirmed this prediction (Figure 2E). This effect was not observed in blood samples (Figure S2) and supports a biological, rather than technical influence on read depth, as technical influences do not discriminate between sexes or chromosomes.

A final strong test of the hypothesis that read depth reflects ongoing DNA replication was provided by a comparison of different cell types: LCLs and embryonic stem cells (ESCs). The DNA replication timing profiles of LCLs and ESCs differ across 20%–30% of the genome (Ryba et al., 2010; Hansen et al., 2010). We sequenced genomic DNA from proliferating ESCs and found that read depth in ESCs matched profiles of

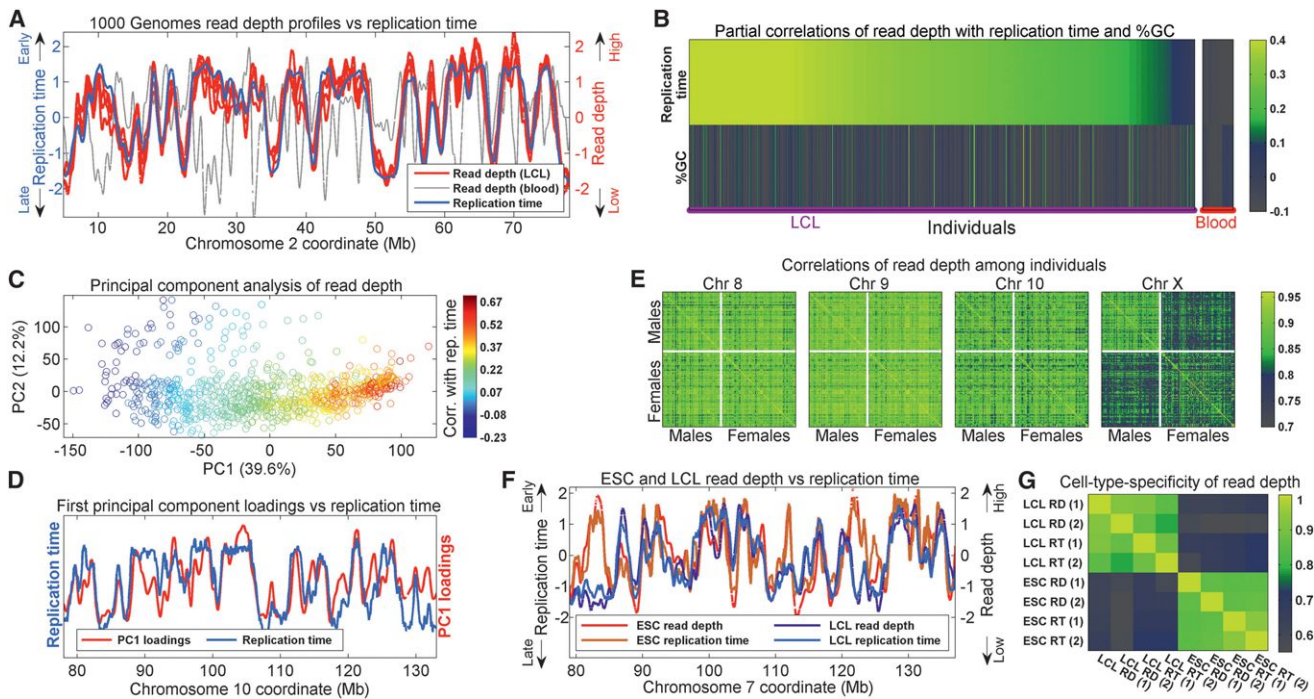


Figure 2. DNA Replication Activity Is Visible in Sequence Data from the 1000 Genomes Project

(A) Long-range fluctuations in read depth along chromosomes follow the DNA replication profile in DNA derived from cultured cells but not in DNA derived from blood. Shown are smoothed, Z normalized read depth profiles of genomic DNA from four 1000 Genomes samples derived from LCLs (red) and one DNA sample derived from blood (gray), along with the LCL replication timing profile (blue).

(B) Read depth is correlated with DNA replication timing to varying extents in different samples (as expected from samples with different proportions of cells in S phase), but is not correlated with GC content. Shown are partial correlations of (unsmoothed) read depth with replication timing (top) and with GC content (bottom), in each case controlling for the other variable (see Figure S1 for complete correlations and sample annotations). Each column corresponds to one of 946 individuals sequenced in the 1000 Genomes Project, sorted by their correlation between read depth and replication timing. Read depth in genomic DNA from blood samples did not correlate with replication timing.

(C) DNA replication timing is the major influence on read depth variation among LCL samples, as determined by principal component analysis. Each circle represents one of 882 LCL samples; color indicates the correlation of read depth with replication timing.

(D) The coefficients (chromosomal loadings) of the first principal component (in D) correspond to the DNA replication timing profile.

(E) A biological signature of the unstructured, “random” replication of inactive X chromosomes from females (Koren and McCarroll, 2014) is apparent in read depth. Inter-individual correlations of read depth along the genome of 161 individuals (see text) are reduced on the X chromosome when comparisons involve a female sample.

(F) Sequencing of DNA from embryonic stem cells (ESCs) identifies ESC-specific replication timing profiles. Shown are read depth profiles of ESCs and LCLs derived from whole-genome sequencing, along with the corresponding S/G1 replication timing profiles. ESC replication timing data is from Ryba et al. (2010). (G) Read depth and replication timing closely track each other within a given cell type (ESC or LCL) and equally distinguish between cell types. Quantitative genome-wide comparison of read depth and replication profiles of ESCs and LCLs (two profiles of each are shown). LCL replication timing is from this study (profile 1) and Ryba et al. (2010) (profile 2). ESC replication timing data is from Ryba et al. (2010). RD, read depth; RT, replication timing. See also Figures S1 and S2 (and Table S1).

replication timing in ESCs and LCLs wherever the profiles were similar between cell types, but matched the ESC profiles wherever ESCs and LCLs had different replication timing profiles (Figures 2F and 2G). Within a cell type, read depth and replication profiles were virtually indistinguishable, whereas many loci exhibited consistent differences between the two cell types that were visible in both whole-genome sequence and explicit replication profiles (Figures 2F and 2G).

DNA Replication Polymorphisms in a Population Cohort

The results described above established the existence of replication timing variation among humans and demonstrate that read depth in whole-genome sequence data from proliferating

cells reflects active DNA replication. These observations raise the intriguing possibility that one could use data from the 1000 Genomes Project to study variation in replication timing within populations and to learn whether it is under genetic control.

We searched for replication variants in the samples sequenced by the 1000 Genomes Project, focusing on 161 DNA samples that appeared to be derived from cultures containing the largest fraction (5%–20%) of S phase cells at the time they were harvested (based on the correlations of read depth fluctuations to replication timing; Figure S1E; Table S2; Extended Experimental Procedures). We excluded genomic regions with evidence of copy number variation (CNV), visible as large-magnitude, stepwise changes in copy number and focused on the

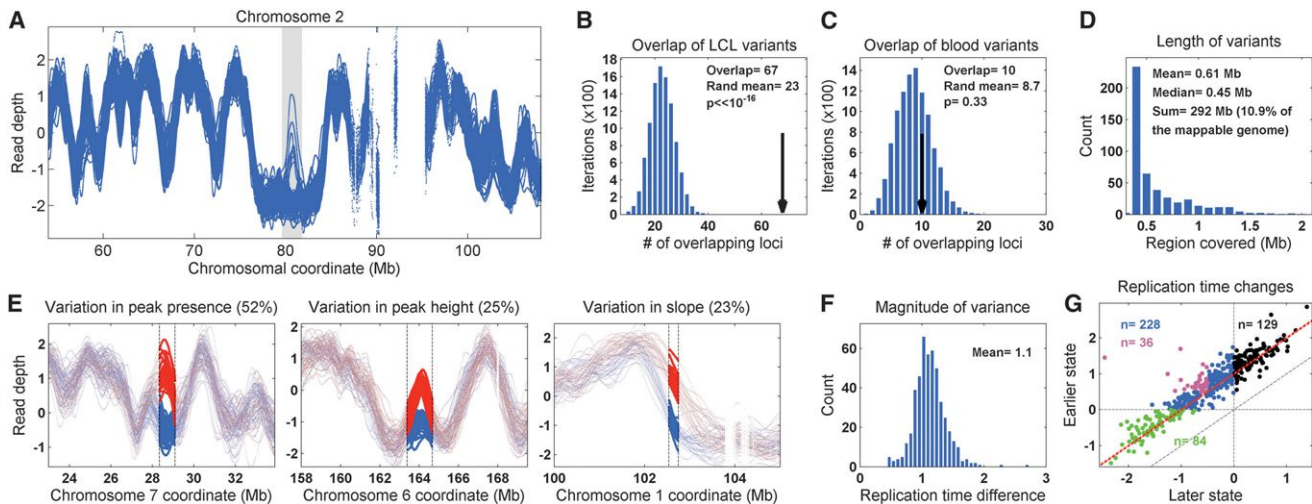


Figure 3. Variation in DNA Replication Timing Is Common in the Human Population

(A) Patterns of read-depth variation among 1000 Genomes individuals indicate the presence of a polymorphic replication origin (gray shaded area). This is the same replication variant shown in Figure 1C as variable in replication timing in the six individuals.

(B) Candidate replication variants identified in the population-based analysis of whole-genome sequence data from the 1000 Genome Project significantly overlap with replication variants identified from direct S/G1 replication profiling of six individuals. Black arrow: number of overlapping variants; blue bars: number of overlapping variants in 10,000 permutations of variant locations.

(C) Loci with the greatest variation in read depth among blood-derived DNA samples from the 1000 Genomes Project did not significantly overlap with variants identified by replication profiling.

(D) Replication variants collectively cover more than 10% of the mappable human genome. Shown is the length distribution of genomic regions affected by replication variants.

(E) Forms of replication variation. The frequency of each variant type is indicated.

(F) The size distribution of replication variants (average replication timing/read depth differences between the early and late replication state in each variant).

(G) Comparison of the replication timing of the early and late states in each individual replication variant locus. Red line: replication difference of 1 SD; black dots: shifts between early and earlier replication; blue dots: shifts between early and late replication (purple dots are loci that shift from under -0.5 to over 0.5 , i.e., the most significant changes between early and late replication); green dots: shifts between late and later replication.

See also Figure S4 and Tables S2 and S3.

lower-amplitude, continuous fluctuations in copy number that reflect active DNA replication (Figure S3; Extended Experimental Procedures).

We identified 361 population variants in the read depth-derived DNA replication timing profiles (Extended Experimental Procedures). Replication variants identified from the 1000 Genomes data significantly ($p < 10^{-16}$) overlapped the replication variants we had identified by direct replication profiling of six individuals (Figures 3A and 3B; as a negative control, loci with the strongest read depth variation across the 64 blood-derived DNA samples exhibited no significant overlap with the replication variants in the six individuals; Figure 3C). To obtain a final set of replication variants, we combined the variants derived from the six individuals with those ascertained from the 1000 Genomes Project individuals and reevaluated the differences among all individuals specifically in these regions (Extended Experimental Procedures). This resulted in a total of 477 variants (Table S3) that spanned 610 kb on average and cumulatively spanned 292 Mb (Figure 3D).

In over 50% of the variants, individuals differed by the presence or absence of a read depth peak, which we interpret as a gain or loss of the activity of a replication origin or origin cluster (Figure 3E). Approximately 25% of replication variants involved quantitative variation in peak height, or the average utilization

or activation timing of a replication origin. The remainder of the variants involved a shift of a replication slope region (transition region; Figure 3E), as could arise if a replication initiation zone was variable in length. Most replication variants were common, with each replication state at each locus shared among multiple individuals (Figures 3A, 3E, and S4); this at least partially reflects our ascertainment approach and does not preclude the possibility of a larger number of rare replication variants that were not detected.

DNA Replication Timing Is Influenced by *cis*-Acting Genetic Variants

The availability of replication timing information for 161 individuals at hundreds of different sites made it possible to search for genetic influences on replication timing. We treated locus-specific replication timing as a quantitative trait (one for each replication variant locus) and analyzed the association of replication timing to sequence variation in the same individuals, using sequence variation data from the 1000 Genomes Project. To reduce the burden of multiple hypothesis testing, we also performed a *cis*-focused association test restricted to genetic variants near each replication variant region (Degner et al., 2012; Lappalainen et al., 2013; Kilpinen et al., 2013; McVicker et al., 2013; Kasowski et al., 2013) (Extended Experimental Procedures).

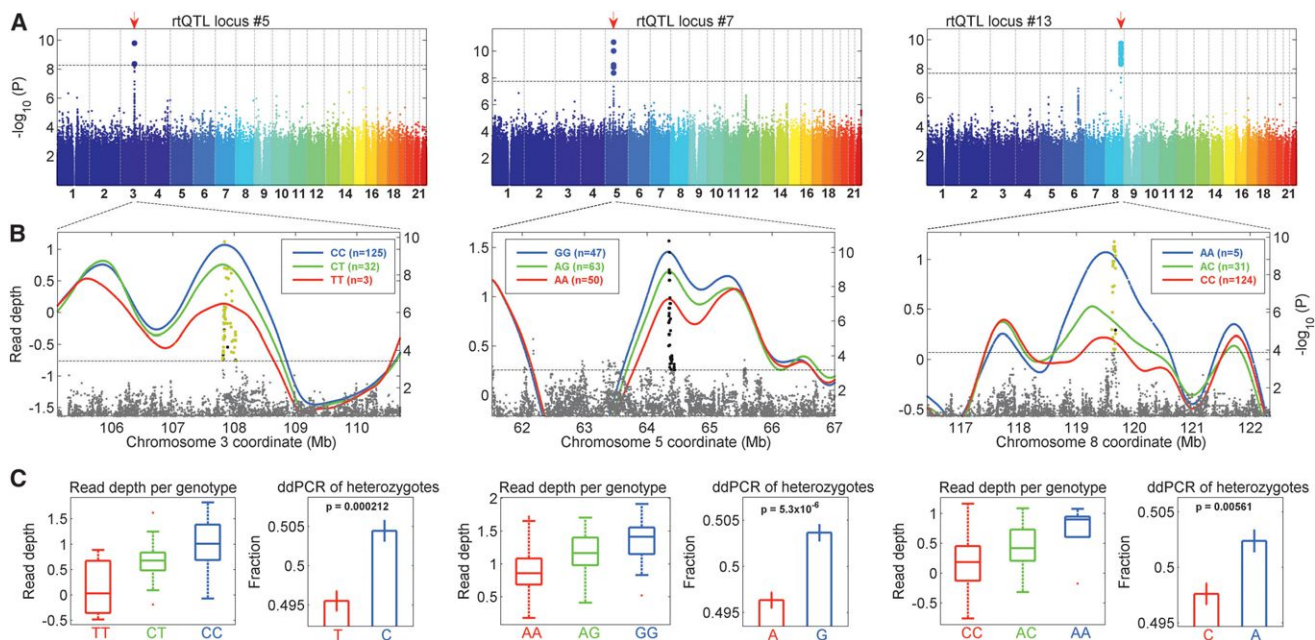


Figure 4. Replication Timing Quantitative Trait Loci

Genetic variants underlie differences in DNA replication timing among individuals. Shown are three examples of replication variants with significant genetic association (additional examples are in Figures 5 and S5).

(A) Variation in replication timing of a specific locus is strongly associated with SNPs that map within the locus itself. Shown are Manhattan plots of genome-wide association of genetic variants with replication timing. Red arrow: genomic location of the tested replication variant region. Black dashed line: genome-wide association significance threshold.

(B) Detailed genetic associations in replication variant regions (dots; right axis) along with replication (read depth) profiles (left axis) for individuals with each of the three genotypes of the most strongly associated SNPs. Yellow dots denote rtQTL SNPs that were also eQTLs for a nearby gene.

(C) Left panels: distribution of read depth for individuals with each of the genotypes of the SNP most strongly associated with each variant. Right panels: droplet digital PCR (ddPCR) analysis confirms that the allele associated with early replication is also overrepresented in genomic DNA from heterozygous individuals, consistent with a *cis*-acting, allele-specific effect on DNA replication timing.

See also Tables S4 and S5.

We identified 20 replication variant loci with significant sequence associations in *cis* (nominal $p = 10^{-5}$ to 10^{-13}), of which eight were identified in the genome-wide scan and an additional twelve in the *cis*-localized scan (Figures 4, 5, and S5; Table S4). As with other genetic traits studied for association with common variants, replication-timing phenotypes tended to associate to haplotypes of many variants in linkage disequilibrium (LD) with one another.

To critically evaluate these associations, we used data from an additional 334 samples from the 1000 Genomes Project; these samples, which had weaker signatures of replication timing ($r = 0.2$ – 0.4), had not been included in the initial scan. At each of the 20 loci, we tested whether the “index SNP” (the most strongly associating SNP) from the initial analysis also associated with measurements of replication timing in the other samples. Despite the lower power to detect replication timing associations using these samples, 16 of the associated loci were replicated with p values of between 10^{-24} and 0.05, all with the same direction of allelic influence as the original samples. The index SNPs at the remaining four loci were not significant in the replication analysis, reflecting an unknown combination of partial power and some false positives in the initial scan (Figure S5).

We also searched for *trans*-associations (associations to variants outside the replication variant region); however, our sample

set is composed of individuals from many different populations, making such an analysis vulnerable to artifacts of population structure. Indeed, the 17 identified putative *trans*-associations did not map to genes related to DNA replication or related pathways and were not considered further (see Extended Experimental Procedures).

We refer to genetic variants that associate with replication timing as replication timing quantitative trait loci (rtQTLs). The 20 rtQTL haplotypes consisted entirely of SNPs and short indels (rather than large structural polymorphisms), indicating that fine-scale sequence variation can be sufficient to affect DNA replication timing on megabase scales (we note that CNVs and other forms of variation could influence replication timing at loci not identified here). The implicated genetic variants were almost always located in the immediate vicinity of a replication timing peak (median distance = 52 kb, $p = 7.2 \times 10^{-5}$) (Figures 4B and 5), suggesting that rtQTLs typically affect DNA replication by affecting replication origins. The implicated rtQTL haplotypes were generally small (2–160 kb, median = 20 kb), yet the regions whose replication timing associated with these haplotypes were 4–600 times larger, encompassing 0.39–1.86 Mb (median = 0.66 Mb) of surrounding sequence (Figures 4 and 5).

Individuals heterozygous for rtQTL SNPs had replication timing phenotypes intermediate between those of homozygous

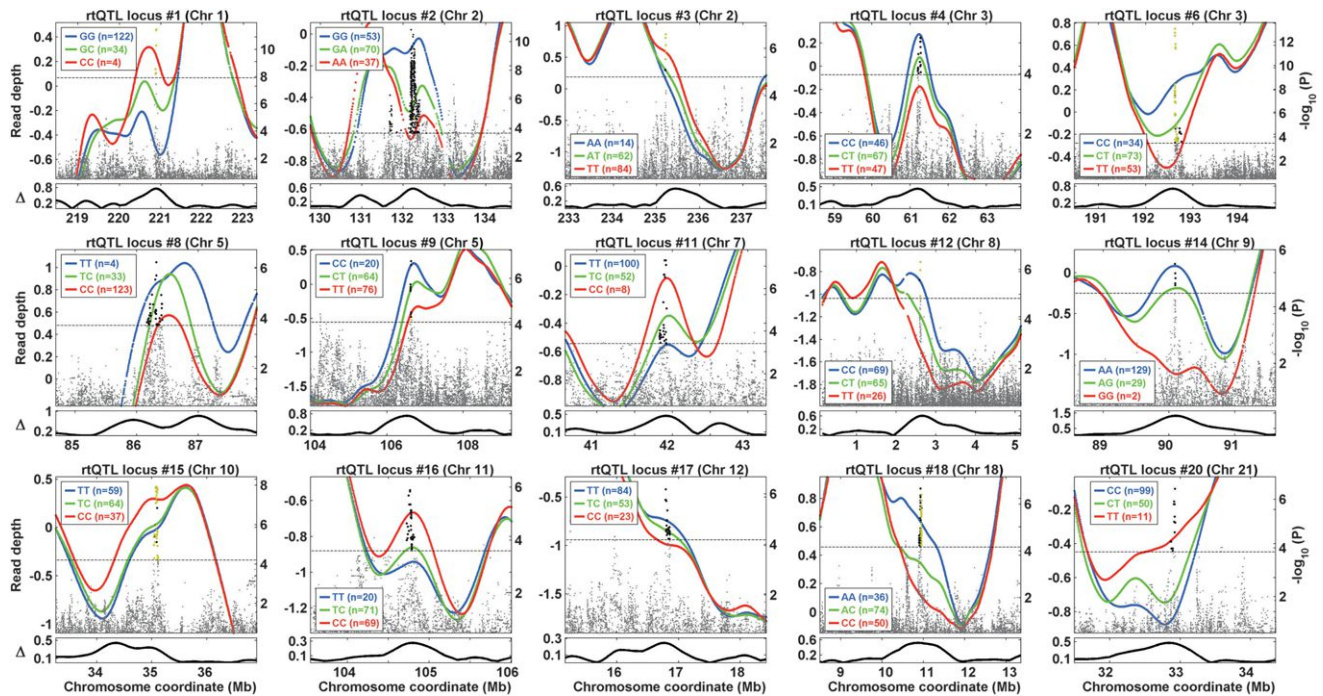


Figure 5. rQTLs Involve Variable Use of Replication Origins and Exert Long-Range Effects on Replication Timing

rQTLs involve associations with sets of markers in the immediate vicinity of replication origins and affect the replication timing of megabases of surrounding DNA. Plots are as in Figure 4B. The lower graphs in each panel (bold black line) show that replication timing differences gradually decrease with distance from rQTL loci. Figure S5 shows a zoomed-in version of all association results, as well as an additional two rQTL loci that were not clearly associated with replication origins. See also Figure S5 and Table S4.

individuals (Figures 4B and 4C). This could be due to having one earlier- and one later-replicating version of the locus on their two chromosomal copies, if rQTLs are due to allele-specific, *cis*-acting influences of DNA sequence on replication timing, as opposed to *trans*-acting or nongenetic effects. Individuals heterozygous for rQTL SNPs should therefore exhibit allelic asynchrony of replication at the rQTL loci and have more copies of the early-replicating allele than the late-replicating allele in their genomic DNA. To test this prediction, we used droplet digital PCR (Hindson et al., 2011) to measure the allelic content of the genomic DNA at four rQTL loci each in LCL-derived DNA from 18–35 heterozygous individuals. At all four loci, the allele associated with earlier replication timing (at a population level) also exhibited greater abundance ($p = 0.005 - 5.3 \times 10^{-6}$) within genomic DNA from heterozygous individuals (Figure 4C), while control SNPs that were not in LD with the rQTL SNPs were not significantly skewed (Extended Experimental Procedures). These results confirm our sequencing- and population-based inference and are consistent with a model in which genetic variation affects replication timing in an allele-specific, *cis*-acting manner.

DNA Replication Is Associated with a Long-Range Effect on Gene Expression Levels

Replication origin activity is associated with open chromatin structure, and DNA replication timing is generally correlated with the levels of gene expression across a genome (Rhind and Gilbert, 2013). We therefore hypothesized that rQTLs may

operate by influencing chromatin states. We compared the locations of rQTLs to the locations of enhancers, defined as DNA segments of ~500 bp containing combinations of histone modifications that promote expression of nearby genes (Ernst et al., 2011). We found a significant enrichment of rQTLs within enhancer regions that were specifically active in LCLs (out of nine cell types examined; Table S6); 11 of 20 rQTL loci contained sequence variants within LCL enhancers, even though the latter cover <1% of the genome (enrichment $\chi^2 p < 10^{-16}$). This relationship suggested that rQTLs may affect DNA replication by promoting an open chromatin structure, prompting us to analyze more closely their relationship to gene expression.

To explore in more detail the relationship between DNA replication timing and gene expression levels at regions implicated by rQTLs, we utilized a recent RNA-seq analysis of gene expression in 462 LCL samples from the 1000 Genomes Project (Lappalainen et al., 2013). We first compared the locations of expression quantitative trait loci (eQTLs) identified in the RNA-seq study with the locations of rQTLs. At nine of the 20 rQTL loci, the implicated SNPs overlapped *cis*-eQTLs (Figures 4B, 5, and S5), even though eQTLs comprised <0.02% of the genome (enrichment $\chi^2 p < 10^{-16}$). Moreover, in eight of those nine cases, the rQTL alleles that associated with early replication were also the alleles associated with elevated expression levels. This observation provides independent confirmation that our rQTL findings, which were made entirely from genomic DNA (without any analysis of RNA), relate to functional aspects of genome

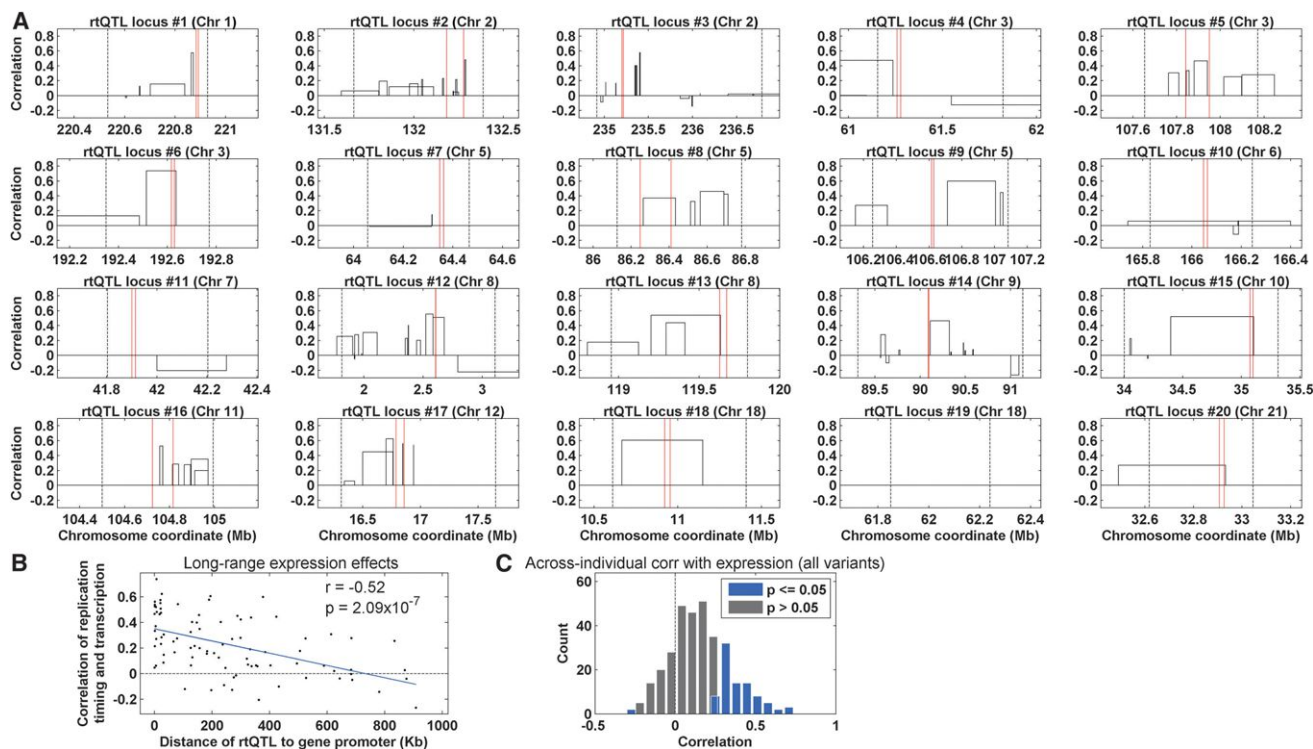


Figure 6. Replication Timing Associates with Gene Expression Levels

Individuals whose genomes exhibit earlier replication at a replication variant locus also tend to exhibit higher average expression of genes across the entire zone of replication.

(A) Correlations between expression levels and replication timing, for the subset of rtQTL loci affecting the replication timing of expressed genes (16 of the 20 rtQTL loci), across 53 individuals, for each gene within the rtQTL-implicated replication variant regions. Dashed black lines: replication variant region borders; red lines: rtQTL association region.

(B) The correlation between replication timing and gene expression decreases as a function of gene distance from the rtQTL SNPs.

(C) The distribution of correlations between replication timing and gene expression across individuals, for all replication variants that contained expressed genes. See also Figures S6 and S7 and Table S6.

biology. Furthermore, the tendency of expressed genes to be in early replicating regions of the genome may reflect shared genetic influences (e.g., influences of genetic variation on open chromatin).

An important distinction between eQTLs and rtQTLs is that most eQTLs directly affect the expression of genes in their immediate vicinity (median distance of 20 kb between SNPs and gene promoter, for the eQTLs overlapping rtQTLs), whereas the rtQTLs associate with the replication timing of megabases of surrounding DNA (median = 660 kb). The order-of-magnitude difference in the scale of the effects of rtQTLs and eQTLs provided a unique opportunity to address a long-standing question—can DNA replication timing itself influence gene expression levels in proliferating cells? We addressed this question by testing for elevated expression of genes across the entire, megabase-scale regions affected by rtQTLs. Focusing on 53 individuals for which both gene expression and replication timing data were available, we compared interindividual variations in replication timing to interindividual variation in gene expression levels in each of the 20 regions implicated by rtQTLs. At each locus, we considered both an aggregate measure of gene expression (across all genes in the replication-affected region) (Figure S6)

as well as the relationship to each individual gene (Figure 6). Individuals with earlier replication of a locus strongly tended to also have higher expression levels of genes throughout the locus (Figure S6), including modest but consistent relationships to expression variation for almost every individual gene (Figure 6). Strikingly, early replication timing consistently correlated with greater gene expression up to distances of ~500 kb, an order of magnitude larger than the typical range of eQTLs, or of the nine eQTLs that overlapped with rtQTLs (Figure 6). These results suggest that replication timing can regulate gene expression levels in proliferating cells and that such effects can be exerted over long genomic distances.

The relationships of early replication to elevated levels of gene expression across individuals also extended to the remainder of the 477 replication timing variants (for which rtQTLs have not currently been identified) (Figure 6C), and replication variant sites were significantly enriched for eQTLs compared to random genomic sites (Figure S7).

Finally, we note that despite the links between DNA replication timing and gene expression, three rtQTL loci were almost completely devoid of transcription (Figure 6A). Thus, while replication timing and gene expression may share some regulatory

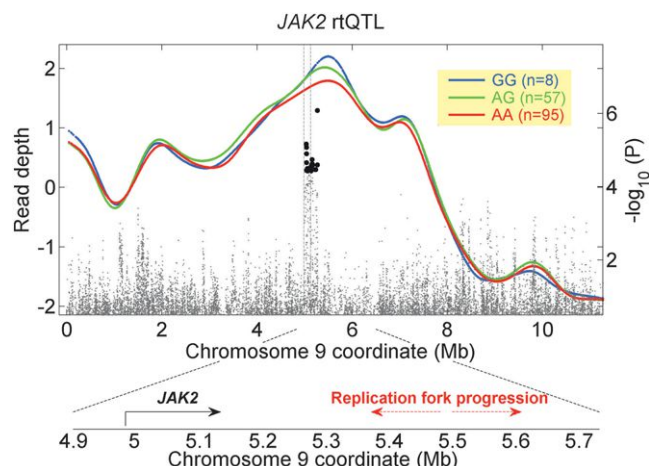


Figure 7. An rtQTL at the *JAK2* Locus

A common allele at a SNP downstream of *JAK2*, previously associated with increased *JAK2* mutation rates, is also associated with very early replication (higher peak) of an adjacent origin in an early replicating fragile site (ERFS) region. *JAK2* (dashed vertical lines) is transcribed toward the inferred replication origin (the peak). The heights of the black points show the level of association of SNPs to the replication timing of this locus, on the scale shown on the right. Diagram on the bottom depicts the location and transcriptional orientation of *JAK2* compared to the direction of replication fork progression from the nearby origin.

influences (such as open chromatin), each process appears to be independently controlled. In particular, transcription is not required for the establishment of rtQTLs.

An rtQTL Links Early Origin Activity to *JAK2* Mutations that Lead to Myeloproliferative Neoplasms

An intriguing implication of rtQTLs is that inherited alleles could modify mutation rates in their genomic vicinity by affecting the replication timing of nearby DNA. A medically important example of polymorphism-associated mutation rates involves the Janus Kinase 2 (*JAK2*) locus. *JAK2* is strongly expressed in blood cells including hematopoietic stem cells, B lymphocytes, and LCLs; because *JAK2* transduces growth signals, activating *JAK2* mutations (e.g., *JAK2*^{V617F}) that arise in individual cells cause clonal expansions that result in myeloproliferative neoplasms and can transform into hematological malignancies. These activating *JAK2* mutations have been shown to arise more frequently in carriers of a “predisposing haplotype” defined by specific alleles at genetic markers across *JAK2* (Olcaydu et al., 2009; Jones et al., 2009; Kilpivaara et al., 2009) and to arise in *cis* with respect to this haplotype (i.e., on the same chromosomal copy) (Olcaydu et al., 2009; Jones et al., 2009; Kilpivaara et al., 2009). The mechanism underlying this relationship is unknown. *JAK2* has also been identified as an early replicating fragile site (ERFS) in B lymphocytes (Barlow et al., 2013). ERFSs are genomic loci at which early origin activation can lead to double strand breaks, particularly in the presence of nearby transcription, with consequently elevated mutation rates at distances of up to hundreds of kilobases from the break site (Barlow et al., 2013; Pedersen and De, 2013; Drier et al., 2013; Jones et al., 2013; Deem et al., 2011; Wang and Vasquez, 2004).

We evaluated the possibility that replication timing variation could explain the mutability of the *JAK2* haplotype and specifically that the mutation-predisposing haplotype is an rtQTL. We found a replication variant near *JAK2*, which was just below the significance threshold of our genome-wide screen for replication variants. Replication at *JAK2* involved an unusually early replicating origin (i.e., a high peak on the replication profile; Figure 7), consistent with the identification of the same locus as an ERFS (Barlow et al., 2013). We also found that the direction of replication fork progression is opposite the direction of *JAK2* transcription (Figure 7), consistent with a model in which chromosome fragility is enhanced by head-on collisions between the replication and transcription machinery. Most importantly, the inherited *JAK2* alleles that predispose to *JAK2* mutations all associated strongly ($p < 4.5 \times 10^{-4}$) with earlier or more efficient activation of the origin (i.e., a higher replication peak; Figure 7) and were among the peak SNPs for the rtQTL (Figure 7). Taken together, these data are consistent with a model in which chromosome fragility, enhanced by interference between the replication and transcription machinery, underlies the mutations in *JAK2* and does so more frequently in individuals in whom replication activity from the origin is earlier and/or more efficient.

DISCUSSION

How eukaryotic genomes specify the timing of replication origin activation is a long-standing mystery. We show here that locus-specific replication timing varies among humans and is influenced by inherited genetic polymorphism. Replication variants involve alterations in the replication timing of large (200 kb–2 Mb) chromosomal regions. Most if not all of these variants relate to differences in replication origin (or origin cluster) activity, as inferred from replication timing peak structures. We discovered SNP haplotypes that associate with DNA replication timing, which we call replication timing QTLs (rtQTLs). The genetic variation implicated at rtQTLs tends to be at or very close to the inferred replication origin. Given the overlap between rtQTLs and enhancers in the same cell type, rtQTLs may affect DNA replication by promoting an open chromatin structure that is permissible for origin firing. Alternatively, some rtQTLs may alter the DNA sequences bound by factors that promote origin firing. Understanding the mode of action of rtQTLs will illuminate the complex process of replication timing control.

To study DNA replication timing, we made use of genome sequence data from the 1000 Genomes Project, which was designed primarily as a study of genome sequence variation (and not a functional study of DNA replication). As a result, the discovery power in the current study was limited by the low read depth (3–5×), the relatively small number of individuals analyzed, and the lack of any deliberate enrichment for S phase cells. Consequently, we have likely found only the rtQTLs with the strongest effect on replication timing and that arise from common alleles. Replication timing is likely shaped by multiple genetic and epigenetic factors and will require more powerful analyses to identify the full sets of underlying factors at each locus. We expect that subsequent work will identify far more rtQTLs in LCLs and other cell types. Identification of a larger number of rtQTLs will facilitate the analysis of their common features and their molecular mode

of action and pave the way for an understanding of the regulation of replication origin activity. Furthermore, identification of the causal variants that control replication origin activity will make it possible to manipulate replication timing experimentally, providing new ways of investigating the causes and consequences of DNA replication timing.

An intriguing implication of our results relates to the relationship between DNA replication timing and the generation of mutations. DNA replication timing is associated with mutation rate variation across the genome in two important ways. First, late-replicating DNA is generally more prone to mutation than early-replicating DNA. Late replication is also associated with increased levels of DNA breakage at common fragile sites (CFSS) (Letessier et al., 2011); notably in this regard, the replication variants we identified overlap 19 CFSSs, including *FRA3B*, the most common fragile site in lymphocytes. Second, elevated mutation rates also occur in regions with high transcriptional activity in the vicinity of early replicating origins due to collisions between the replication and transcription machineries, which lead to chromosome fragility, double strand breaks, single-stranded DNA (ssDNA) formation, and error-prone DNA synthesis (Barlow et al., 2013; Pedersen and De, 2013; Drier et al., 2013; Jones et al., 2013; Deem et al., 2011; Wang and Vasquez, 2004). Genetic variants that affect DNA replication timing therefore have the potential to affect mutation rates in their vicinity. Such an effect would have important implications for evolution and for disease. First, rtQTL alleles conferring regional late replication or early origin activity in the vicinity of active genes could function as *cis*-acting mutators that cosegregate, via genetic linkage, with the mutations they induce, providing a mechanism for evolutionary optimization of local mutation rates in sexual species (Martincorena and Luscombe, 2013). Second, rtQTLs may serve as common, inherited genetic polymorphisms that affect the probability of somatic mutation at specific loci. Diseases with high heritability are often assumed to be distinct from diseases of somatic mutation. Our results suggest, however, that inherited polymorphism can consign a genomic region to late replication or create early replicating fragile sites in a particular tissue, thereby increasing the likelihood that it will acquire somatic mutations in that tissue. At the *JAK2* kinase locus, for example, the same SNP haplotype is associated with both early origin activation and elevated mutation rates that can lead to myeloproliferative neoplasms. Altered replication timing in a relevant cell population could thus be a means by which inherited variation influences somatic mutation rates and consequentially, disease and cancer susceptibility.

The presence of a substantial subpopulation of S phase cells in expanding cell cultures appears to endow whole-genome sequences derived from such samples with information about ongoing DNA replication activity. The influence of DNA replication is directly related to the proportion of cells that are in S phase, which for cultured cells depends on their growth phase: exponentially growing cultures will contain the largest fraction of replicating cells, while quiescent cultures will tend to contain mostly cells in G1 phase. Replication timing could influence any measurement of DNA content (array- or sequencing-based) that has been made from proliferating cells (e.g., studies of copy number variation and chromatin states). Copy number detection

in single cells, for example during preimplantation genetic diagnosis (PGD), is also more prone to false CNV detection when a cell is in S phase (Dimitriadou et al., 2014). Replication timing will need to be carefully considered as a potential confounding variable in genomic studies. On the other hand, the sequencing of genomic DNA derived from proliferating cells could become a routine way of studying replication dynamics. This approach will enable the study of DNA replication dynamics in a wide range of experimental conditions, cell types, and species, in a technically straightforward way.

EXPERIMENTAL PROCEDURES

Replication variants were discovered in replication timing data of six individuals (Koren et al., 2012) by pairwise comparisons of consecutive 200 kb windows along the genome, selection of windows with a p value $<10^{-10}$ (t test), and consolidation of significant windows within 200 kb of other significant windows into discrete variant loci. Read depth measurements in 10 kb windows from samples from the 1000 Genomes Project (1000 Genome Project Consortium, 2012; Handsaker et al., 2011) were compared to replication timing profiles; for the 161 samples with a genome-wide correlation of >0.4 between read depth and replication timing, replication variants were identified as above and the two lists of replication variants were consolidated into a total of 477 replication variant loci. At each locus, quantitative measurements of replication timing were derived from the 1000 Genomes data across the 161 individuals and were correlated with the genotypes (from the 1000 Genomes Project) of these same individuals. One thousand permutations of sample genotypes were performed in order to obtain an empirical significance threshold for associations with genetic variants. We performed a genome wide association test with over 7.5 million genetic variants with an allele frequency >0.05 in the tested individuals; as well as a *cis*-focused association test with genetic variants only within each replication variant locus. rtQTLs were validated using droplet digital PCR (Hindson et al., 2011) with allele-specific probes. Expression data and eQTLs were from Lappalainen et al. (2013). The two human embryonic stem cell lines used for this study, HUES64 and HUES63, were donated for research following informed consent under protocols reviewed and approved by the Committee on the Use of Human Subjects (IRB) and the Embryonic Stem Cell Research Oversight Committee (ESCRO) at Harvard University. See Extended Experimental Procedures for further details.

ACCESSION NUMBERS

The database of Genotypes and Phenotypes (dbGaP) accession number for the ESC genome sequences reported in this paper is phs000825.v1.p1.

SUPPLEMENTAL INFORMATION

Supplemental Information includes Extended Experimental Procedures, seven figures, and six tables and can be found with this article online at <http://dx.doi.org/10.1016/j.cell.2014.10.025>.

ACKNOWLEDGMENTS

We thank Vanessa Van Doren for technical assistance and David Altshuler, Chris Patil, Giulio Genovese, Sam Rose, and Itamar Simon for discussions and comments on the manuscript. This work was supported by the National Human Genome Research Institute (R01 HG 006855 to S.A.M.), the Integrative Life Seventh Framework Programme (grant 315997 to R.K.), the Stanley Center for Psychiatric Research, the Howard Hughes Medical Institute, and the Harvard Stem Cell Institute.

Received: June 8, 2014

Revised: September 2, 2014

Accepted: October 8, 2014

Published: November 13, 2014

REFERENCES

- Aird, D., Ross, M.G., Chen, W.S., Danielsson, M., Fennell, T., Russ, C., Jaffe, D.B., Nusbaum, C., and Gnirke, A. (2011). Analyzing and minimizing PCR amplification bias in Illumina sequencing libraries. *Genome Biol.* 12, R18.
- Barlow, J.H., Faryabi, R.B., Callén, E., Wong, N., Malhowski, A., Chen, H.T., Gutierrez-Cruz, G., Sun, H.-W., McKinnon, P., Wright, G., et al. (2013). Identification of early replicating fragile sites that contribute to genome instability. *Cell* 152, 620–632.
- Deem, A., Keszthelyi, A., Blackgrove, T., Vayl, A., Coffey, B., Mathur, R., Chabes, A., and Malkova, A. (2011). Break-induced replication is highly inaccurate. *PLoS Biol.* 9, e1000594.
- Degner, J.F., Pai, A.A., Pique-Regi, R., Veyrieras, J.-B., Gaffney, D.J., Pickrell, J.K., De Leon, S., Michelini, K., Lewellen, N., Crawford, G.E., et al. (2012). DNase I sensitivity QTLs are a major determinant of human expression variation. *Nature* 482, 390–394.
- Dimitriadou, E., Van der Aa, N., Cheng, J., Voet, T., and Vermeesch, J.R. (2014). Single cell segmental aneuploidy detection is compromised by S phase. *Mol. Cytogenet.* 7, 46.
- Diskin, S.J., Li, M., Hou, C., Yang, S., Glessner, J., Hakonarson, H., Bucan, M., Maris, J.M., and Wang, K. (2008). Adjustment of genomic waves in signal intensities from whole-genome SNP genotyping platforms. *Nucleic Acids Res.* 36, e126.
- Drier, Y., Lawrence, M.S., Carter, S.L., Stewart, C., Gabriel, S.B., Lander, E.S., Meyerson, M., Beroukhi, R., and Getz, G. (2013). Somatic rearrangements across cancer reveal classes of samples with distinct patterns of DNA breakage and rearrangement-induced hypermutability. *Genome Res.* 23, 228–235.
- Ernst, J., Kheradpour, P., Mikkelsen, T.S., Shores, N., Ward, L.D., Epstein, C.B., Zhang, X., Wang, L., Issner, R., Coyne, M., et al. (2011). Mapping and analysis of chromatin state dynamics in nine human cell types. *Nature* 473, 43–49.
- Handsaker, R.E., Korn, J.M., Nemesh, J., and McCarroll, S.A. (2011). Discovery and genotyping of genome structural polymorphism by sequencing on a population scale. *Nat. Genet.* 43, 269–276.
- Hansen, R.S., Thomas, S., Sandstrom, R., Canfield, T.K., Thurman, R.E., Weaver, M., Dorschner, M.O., Gattler, S.M., and Stamatoyannopoulos, J.A. (2010). Sequencing newly replicated DNA reveals widespread plasticity in human replication timing. *Proc. Natl. Acad. Sci. USA* 107, 139–144.
- Hawkins, M., Retkute, R., Müller, C.A., Saner, N., Tanaka, T.U., de Moura, A.P.S., and Nieduszynski, C.A. (2013). High-resolution replication profiles define the stochastic nature of genome replication initiation and termination. *Cell Reports* 5, 1132–1141.
- Hindson, B.J., Ness, K.D., Masquelier, D.A., Belgrader, P., Heredia, N.J., Makarewicz, A.J., Bright, I.J., Lucero, M.Y., Hiddessen, A.L., Legler, T.C., et al. (2011). High-throughput droplet digital PCR system for absolute quantitation of DNA copy number. *Anal. Chem.* 83, 8604–8610.
- Hiratani, I., Ryba, T., Itoh, M., Yokochi, T., Schwaiger, M., Chang, C.-W., Lyou, Y., Townes, T.M., Schübeler, D., and Gilbert, D.M. (2008). Global reorganization of replication domains during embryonic stem cell differentiation. *PLoS Biol.* 6, e245.
- Jones, A.V., Chase, A., Silver, R.T., Oscier, D., Zoi, K., Wang, Y.L., Cario, H., Pahl, H.L., Collins, A., Reiter, A., et al. (2009). JAK2 haplotype is a major risk factor for the development of myeloproliferative neoplasms. *Nat. Genet.* 41, 446–449.
- Jones, R.M., Mortusewicz, O., Afzal, I., Lorvellec, M., García, P., Helleday, T., and Petermann, E. (2013). Increased replication initiation and conflicts with transcription underlie Cyclin E-induced replication stress. *Oncogene* 32, 3744–3753.
- Kasowski, M., Kyriazopoulou-Panagiotopoulou, S., Grubert, F., Zaugg, J.B., Kundaje, A., Liu, Y., Boyle, A.P., Zhang, Q.C., Zakharia, F., Spacek, D.V., et al. (2013). Extensive variation in chromatin states across humans. *Science* 342, 750–752.
- Kilpinen, H., Waszak, S.M., Gschwind, A.R., Raghav, S.K., Witwicki, R.M., Orioli, A., Migliaiavacca, E., Wiederkehr, M., Gutierrez-Arcelus, M., Panousis, N.I., et al. (2013). Coordinated effects of sequence variation on DNA binding, chromatin structure, and transcription. *Science* 342, 744–747.
- Kilpivaara, O., Mukherjee, S., Schram, A.M., Wadleigh, M., Mullally, A., Ebert, B.L., Bass, A., Marubayashi, S., Heguy, A., Garcia-Manero, G., et al. (2009). A germline JAK2 SNP is associated with predisposition to the development of JAK2(V617F)-positive myeloproliferative neoplasms. *Nat. Genet.* 41, 455–459.
- Koren, A., and McCarroll, S.A. (2014). Random replication of the inactive X chromosome. *Genome Res.* 24, 64–69.
- Koren, A., Polak, P., Nemesh, J., Michaelson, J.J., Sebat, J., Sunyaev, S.R., and McCarroll, S.A. (2012). Differential relationship of DNA replication timing to different forms of human mutation and variation. *Am. J. Hum. Genet.* 91, 1033–1040.
- Lappalainen, T., Sammeth, M., Friedländer, M.R., 't Hoen, P.A.C., Monlong, J., Rivas, M.A., González-Porta, M., Kurbatova, N., Griebel, T., Ferreira, P.G., et al.; Geuvadis Consortium (2013). Transcriptome and genome sequencing uncovers functional variation in humans. *Nature* 501, 506–511.
- Lawrence, M.S., Stojanov, P., Polak, P., Kryukov, G.V., Cibulskis, K., Sivachenko, A., Carter, S.L., Stewart, C., Mermel, C.H., Roberts, S.A., et al. (2013). Mutational heterogeneity in cancer and the search for new cancer-associated genes. *Nature* 499, 214–218.
- Leprêtre, F., Villenet, C., Quief, S., Nibourel, O., Jacquemin, C., Troussard, X., Jardin, F., Gibson, F., Kerckaert, J.P., Roumier, C., and Figeac, M. (2010). Waved aCGH: to smooth or not to smooth. *Nucleic Acids Res.* 38, e94.
- Letessier, A., Millot, G.A., Koundrioukoff, S., Lachagès, A.-M., Vogt, N., Hansen, R.S., Maffoy, B., Brison, O., and Debatisse, M. (2011). Cell-type-specific replication initiation programs set fragility of the FRA3B fragile site. *Nature* 470, 120–123.
- Manukjan, G., Tauscher, M., and Steinemann, D. (2013). Replication timing influences DNA copy number determination by array-CGH. *Biotechniques* 55, 231–232.
- Marioni, J.C., Thorne, N.P., Valsesia, A., Fitzgerald, T., Redon, R., Fiegler, H., Andrews, T.D., Stranger, B.E., Lynch, A.G., Dermizakis, E.T., et al. (2007). Breaking the waves: improved detection of copy number variation from microarray-based comparative genomic hybridization. *Genome Biol.* 8, R228.
- Martincorena, I., and Luscombe, N.M. (2013). Non-random mutation: the evolution of targeted hypermutation and hypomutation. *BioEssays* 35, 123–130.
- McVicker, G., van de Geijn, B., Degner, J.F., Cain, C.E., Banovich, N.E., Raj, A., Lewellen, N., Myrthil, M., Gilad, Y., and Pritchard, J.K. (2013). Identification of genetic variants that affect histone modifications in human cells. *Science* 342, 747–749.
- Mukhopadhyay, R., Lajugie, J., Fourel, N., Selzer, A., Schizas, M., Bartholdy, B., Mar, J., Lin, C.M., Martin, M.M., Ryan, M., et al. (2014). Allele-specific genome-wide profiling in human primary erythroblasts reveal replication program organization. *PLoS Genet.* 10, e1004319.
- Olcaydu, D., Harutyunyan, A., Jäger, R., Berg, T., Gisslinger, B., Pabinger, I., Gisslinger, H., and Kralovics, R. (2009). A common JAK2 haplotype confers susceptibility to myeloproliferative neoplasms. *Nat. Genet.* 41, 450–454.
- 1000 Genome Project Consortium (2012). An integrated map of genetic variation from 1,092 human genomes. *Nature* 491, 56–65.
- Patterson, N., Price, A.L., and Reich, D. (2006). Population structure and eigenanalysis. *PLoS Genet.* 2, e190.
- Pedersen, B.S., and De, S. (2013). Loss of heterozygosity preferentially occurs in early replicating regions in cancer genomes. *Nucleic Acids Res.* 41, 7615–7624.
- Pope, B.D., Tsumagari, K., Battaglia, D., Ryba, T., Hiratani, I., Ehrlich, M., and Gilbert, D.M. (2011). DNA replication timing is maintained genome-wide in primary human myoblasts independent of D4Z4 contraction in FSH muscular dystrophy. *PLoS ONE* 6, e27413.
- Raghuraman, M.K., Winzler, E.A., Collingwood, D., Hunt, S., Wodicka, L., Conway, A., Lockhart, D.J., Davis, R.W., Brewer, B.J., and Fangman, W.L. (2001). Replication dynamics of the yeast genome. *Science* 294, 115–121.

- Rhind, N., and Gilbert, D.M. (2013). DNA replication timing. *Cold Spring Harb. Perspect. Biol.* 5, a010132.
- Rocha, E.P. (2008). The organization of the bacterial genome. *Annu. Rev. Genet.* 42, 211–233.
- Ryba, T., Hiratani, I., Lu, J., Itoh, M., Kulik, M., Zhang, J., Schulz, T.C., Robins, A.J., Dalton, S., and Gilbert, D.M. (2010). Evolutionarily conserved replication timing profiles predict long-range chromatin interactions and distinguish closely related cell types. *Genome Res.* 20, 761–770.
- Ryba, T., Battaglia, D., Chang, B.H., Shirley, J.W., Buckley, Q., Pope, B.D., Devadas, M., Druker, B.J., and Gilbert, D.M. (2012). Abnormal developmental control of replication-timing domains in pediatric acute lymphoblastic leukemia. *Genome Res.* 22, 1833–1844.
- Sharp, P.M., Shields, D.C., Wolfe, K.H., and Li, W.H. (1989). Chromosomal location and evolutionary rate variation in enterobacterial genes. *Science* 246, 808–810.
- Slager, J., Kjos, M., Attaiech, L., and Veening, J.-W. (2014). Antibiotic-induced replication stress triggers bacterial competence by increasing gene dosage near the origin. *Cell* 157, 395–406.
- Stamatoyannopoulos, J.A., Adzhubei, I., Thurman, R.E., Kryukov, G.V., Mirkin, S.M., and Sunyaev, S.R. (2009). Human mutation rate associated with DNA replication timing. *Nat. Genet.* 41, 393–395.
- van de Wiel, M.A., Brosens, R., Eilers, P.H.C., Kumps, C., Meijer, G.A., Menten, B., Sistermans, E., Speleman, F., Timmerman, M.E., and Ylstra, B. (2009). Smoothing waves in array CGH tumor profiles. *Bioinformatics* 25, 1099–1104.
- van Heesch, S., Mokry, M., Boskova, V., Junker, W., Mehon, R., Toonen, P., de Bruijn, E., Shull, J.D., Aitman, T.J., Cuppen, E., and Guryev, V. (2013). Systematic biases in DNA copy number originate from isolation procedures. *Genome Biol.* 14, R33.
- Wang, G., and Vasquez, K.M. (2004). Naturally occurring H-DNA-forming sequences are mutagenic in mammalian cells. *Proc. Natl. Acad. Sci. USA* 101, 13448–13453.

ER Contact Sites Define the Position and Timing of Endosome Fission

Ashley A. Rowland,^{1,2} Patrick J. Chitwood,^{1,2} Melissa J. Phillips,¹ and Gia K. Voeltz^{1,*}

¹Department of Molecular, Cellular, and Developmental Biology, University of Colorado, Boulder, CO 80309, USA

²Co-first author

*Correspondence: gia.voeltz@colorado.edu

<http://dx.doi.org/10.1016/j.cell.2014.10.023>

SUMMARY

Endocytic cargo and Rab GTPases are segregated to distinct domains of an endosome. These domains maintain their identity until they undergo fission to traffic cargo. It is not fully understood how segregation of cargo or Rab proteins is maintained along the continuous endosomal membrane or what machinery is required for fission. Endosomes form contact sites with the endoplasmic reticulum (ER) that are maintained during trafficking. Here, we show that stable contacts form between the ER and endosome at constricted sorting domains, and free diffusion of cargo is limited at these positions. We demonstrate that the site of constriction and fission for early and late endosomes is spatially and temporally linked to contact sites with the ER. Lastly, we show that altering ER structure and dynamics reduces the efficiency of endosome fission. Together, these data reveal a surprising role for ER contact in defining the timing and position of endosome fission.

INTRODUCTION

The endocytic pathway is used to internalize components present on the plasma membrane and in the extracellular fluid. After internalization, endocytosed cargo is sorted at multiple steps during trafficking. Cargo destined for degradation at the lysosome is sorted away from both recycled cargo destined for the plasma membrane and cargo trafficked to the Golgi. The sorting of these cargoes must occur prior to endosome fission, and multiple machineries and mechanisms have been identified that contribute to this process (Hanyaloglu and von Zastrow, 2008; Maxfield and McGraw, 2004; Seaman, 2008). Microtubules and their motor proteins, branched actin networks generated by the Arp2/3 activator WASH, the retromer, and structural membrane shaping proteins such as sorting nexins (SNX), have all been implicated in endosome structure and cargo sorting (Gautreau et al., 2014; Hunt et al., 2013; Puthenveedu et al., 2010). However, it is not known what regulates the timing and position of membrane fission to separate the sorted compartments.

Functional contact sites have been observed between the endoplasmic reticulum (ER) and endosomes (Alpy et al., 2013;

Eden et al., 2010; Rocha et al., 2009). Measurements by electron microscopy and tomography have revealed that contact sites between the ER network and individual endosomes exist at multiple discrete positions around the endosome, which additively covers only ~5% of the endosome surface area (Alpy et al., 2013; Friedman et al., 2013). Despite the abundance and discrete nature of these contacts, they appear to be tightly coupled because the two organelles maintain contact even as they are trafficked on the microtubule network (Friedman et al., 2013; Zajac et al., 2013). Endosomes become bound to the ER early in their biogenesis and this association increases with maturation: we found >99% of late endosomes are tightly associated with the ER as they traffic, in contrast to ~50% of early endosomes (Friedman et al., 2013). Thus, ER contact could regulate the biogenesis of endosomes or become targeted to endosomes following a maturation step. Once established, ER contact with endosomes is often maintained despite trafficking, and this suggests important functions occur at the interface.

Two functions have been demonstrated to occur at the ER-endosome interface (van der Kant and Neefjes, 2014). Interactions between (VAMP)-associated protein A (VAP-A) on the ER and the endosome localized partners are thought to regulate cholesterol sensing and lipid transfer. For example, the endosomal protein ORP1L interacts with VAP-A under low cholesterol conditions that could allow for cholesterol exchange, thereby acting as a sensor (Rocha et al., 2009). ER-endosome contact also occurs via the ER-localized phosphatase, PTP1B, which interacts with epidermal growth factor receptor (EGFR), dephosphorylating it to promote incorporation into intraluminal vesicles, a necessary step for EGFR degradation by the lysosome (Eden et al., 2010). Several recent papers also suggest that late endosomes may take up Ca^{2+} from ER stores during their maturation process, however, it remains to be determined whether Ca^{2+} is directly transferred at the interface (Kilpatrick et al., 2013; López-Sanjurjo et al., 2013; Morgan et al., 2013).

The ER also forms contacts with several other organelles (Helle et al., 2013), and its role at these various sites may be analogous. At mitochondria and the plasma membrane, the ER provides Ca^{2+} in different functional contexts (Elbaz and Schuldiner, 2011). Emerging evidence also shows lipids are modified or transferred at the ER interface (Stefan et al., 2013; Toulmay and Prinz, 2011). Recently, we discovered that ER tubules circumscribe mitochondrial constrictions and define the position of mitochondrial fission (Friedman et al., 2011). We predicted that mechanisms of membrane fission may also be conserved between various organelles. Here, we hypothesized and tested

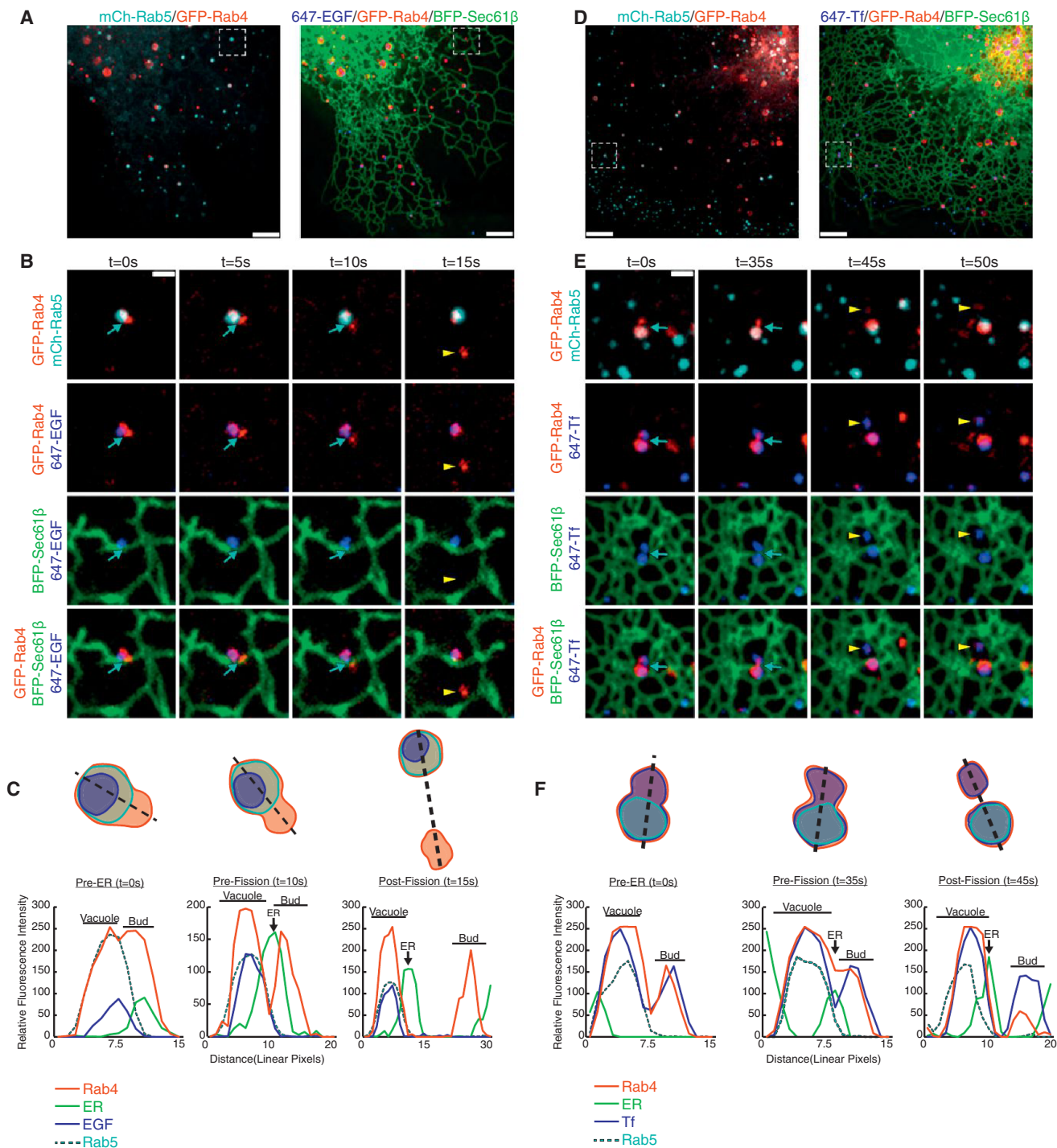


Figure 1. Early Endosome Fission Occurs at ER Contact Sites

(A) Merged images of a live Cos-7 cell expressing mCh-Rab5 (EE in cyan), GFP-Rab4 (EE in red), and BFP-Sec61β (ER in green) that is pulse-labeled with EGF conjugated to Alexa Fluor 647 (cargo in blue).

(B) Magnified time-lapse images of the region boxed in (A) shows an example of early endosome fission. Merged images show the relative location of Rab4, Rab5, ER, and EGF, as indicated over time. The exiting Rab4⁺ bud is marked by a yellow arrowhead. See also Movie S1.

(C) Trace outline of the endosome shown in (B) and the corresponding line-scan analysis of the relative fluorescence intensity (FI) of Rab4, Rab5, ER and EGF for time points: t = 0 s (Pre-ER), t = 10 s (Pre-Fission), and t = 15 s (Post-Fission). Images in (B) and line-scan in (C) reveal that a dynamic ER tubule (marked by a blue arrow) is recruited to the divide between Rab4 compartments just before fission (at t = 10 s).

(D) Merged images taken of a live Cos-7 cell transfected as in (A) that is instead pulse-labeled with Tf conjugated to Alexa Fluor 647 (cargo in blue).

(legend continued on next page)

whether ER contacts define the timing and the position of endosome fission. To test this, we visualized Rab partitioning, cargo sorting, and endosome fission relative to the ER network in live Cos-7 cells.

RESULTS

Dynamic ER Tubules Make Contact with Early Endosome Fission Sites

To visualize early endosome sorting and fission events relative to the ER, cells were cotransfected with GFP-Rab4, mCh-Rab5, and with the general ER marker BFP-Sec61 β and were imaged by live confocal fluorescence microscopy (Figures 1A and 1D). Rab4 localizes to both the vacuolar and budding compartment of the early endosome (Sönnichsen et al., 2000), while Rab5 is exclusively localized to the vacuolar compartment (Sönnichsen et al., 2000; Trischler et al., 1999). Partitioning of cargo also occurs within the early endosome sorting domains. To ensure that the endosomes we visualized were functional, cells were pulse-labeled with epidermal growth factor (EGF) (Figures 1A–1C) or transferrin (Tf) (Figures 1D–1F) cargo conjugated to Alexa Fluor 647. The Rab4 and Rab5 endosomes contain labeled cargo, and their localization and morphology is consistent with previous reports (Sönnichsen et al., 2000; Trischler et al., 1999) (Figures 1A and 1D).

Upon fission, a smaller Rab4⁺ compartment buds off of a larger Rab4⁺/Rab5⁺ vacuolar compartment (Figures 1B and E, top panels; Movies S1 and S2 available online). The fluorescent EGF is confined to the vacuolar Rab4⁺/Rab5⁺ compartment and does not enter into the bud (Figure 1B, panel row 3, yellow arrowhead; Movie S1). In contrast, Tf was not confined to a compartment and was found in the Rab4⁺ bud as well as the Rab4⁺/Rab5⁺ compartment (Figure 1E, panel row 3, yellow arrowhead; Movie S2). Live cell imaging (Figures 1B and 1E) and the corresponding line-scan analysis (Figures 1C and 1F) reveals that in both examples, an ER tubule moves into position at the divide between the two preformed compartments just before fission occurs (compare the position of the ER tubule relative to the two sorting compartments in the Pre-ER and Pre-Fission images and line-scans). These data reveal that ER tubules mark the position of early endosome fission and that there is a temporal relationship between ER recruitment and fission.

ER Tubules Contact Sorting Domains on Early Endosomes

It is difficult to quantitate whether the ER marks all early endosome fission events because early endosomes are small and fission is rapid, hindering our ability to capture instances of clear constriction followed by fission. Thus, we sought ways to slow fission and/or inhibit it at intermediate stages to better resolve the relationship between ER tubules and the position of early endosome constriction and fission. This was accomplished by treating cells with dynasore, a drug which has been shown to

inhibit Dynamin-1, Dynamin-2, and Drp1 *in vitro* (Kirchhausen et al., 2008; Macia et al., 2006). As previously reported (Derivery et al., 2009; Mesaki et al., 2011), Cos-7 cells treated with dynasore accumulate tubular early endosomes (Figure 2A). This phenotype could be a consequence of dynamin inhibition or it could be due to off-target effects, which have been reported (Park et al., 2013). However, our data demonstrate that dynasore treatment is a very useful tool to slow the process of endosome fission and image bona fide fission intermediates at ER contact sites.

Cells were cotransfected with markers of early endosomes (mCh-Rab5) and the ER (GFP-Sec61 β) and were treated with dynasore just prior to imaging. Treatment led to the formation of elongated Rab5⁺ endosomes and did not alter the structure of the ER (Figure 2A) (Derivery et al., 2009; Mesaki et al., 2011). Like untreated endosomes, dynasore-induced tubular endosomes contain both degraded and recycled cargo that partition as expected (Figures S1A and S1B). To further test if tubular endosomes had features characteristic of native endosomes, we investigated whether they contained domains marked by the WASH complex (Figure 2). WASH is a multiprotein complex that activates Arp2/3-mediated actin nucleation at sorting domains (Derivery et al., 2009; Gomez and Billadeau, 2009; Harbour et al., 2012; Jia et al., 2010), and this actin nucleation may participate in cargo sorting (Puthenveedu et al., 2010), or generate force to catalyze fission. To mark the complex, we tagged FAM21, which has been shown to localize to early endosomes, bind the retromer complex, and regulate recruitment of WASH to endosomal sorting domains (Seaman et al., 2013). Cells were cotransfected with GFP-Rab5, BFP-Sec61 β , and mCh-FAM21. In dynasore-treated cells, FAM21 localized to patches and punctate structures along tubular endosomes (Figure 2B). Line-scan analysis demonstrated that FAM21-labeled structures corresponded to positions where Rab5 fluorescence was minimal, indicative of a constriction (see arrows in Figure 2C and 2D). ER tubules also crossed over the positions of FAM21-marked constrictions (see arrows in Figure 2C and 2D). These data demonstrate that sorting domains localize to dynasore-induced elongated endosomes and ER tubule crossings are spatially linked to these domains.

ER Tubules Contact Endosomes at a Barrier to Cargo Diffusion

Our data suggest that dynasore treatment can stall fission at ER-marked endosome sorting domains, and line-scan analysis indicates endosomes are constricted at ER contact sites (Figure 2D). To further establish the importance of the ER at stalled constrictions, we reasoned a functional constriction has the potential to impede the diffusion of cargo. According to the bulk flow hypothesis, membrane-bound cargo that is recycled to the plasma membrane, like transferrin receptor (TfR), diffuses freely in the membrane (Dunn et al., 1989; Mayor et al., 1993). We designed photobleaching techniques to test whether a diffusion barrier

(E) Zoom from (D) shows the relative localization of Rab4, Rab5, ER, and Tf as indicated over time. An ER tubule (marked by blue arrow) is again recruited to the divide between Rab4 compartments just before fission ($t = 35$ s). An exiting Rab4⁺/Tf⁺ bud is marked by a yellow arrowhead. See also Movie S2.

(F) Line-scan analysis of relative FI depicts the position of a dynamic ER tubule to the position and timing of endosome fission for time points: $t = 0$ s (Pre-eR), $t = 35$ s (Pre-Fission), and $t = 45$ s (Post-Fission). Scale bars represent 5 μ m in (A) and (D) and 1 μ m in (B) and (E). s, seconds.

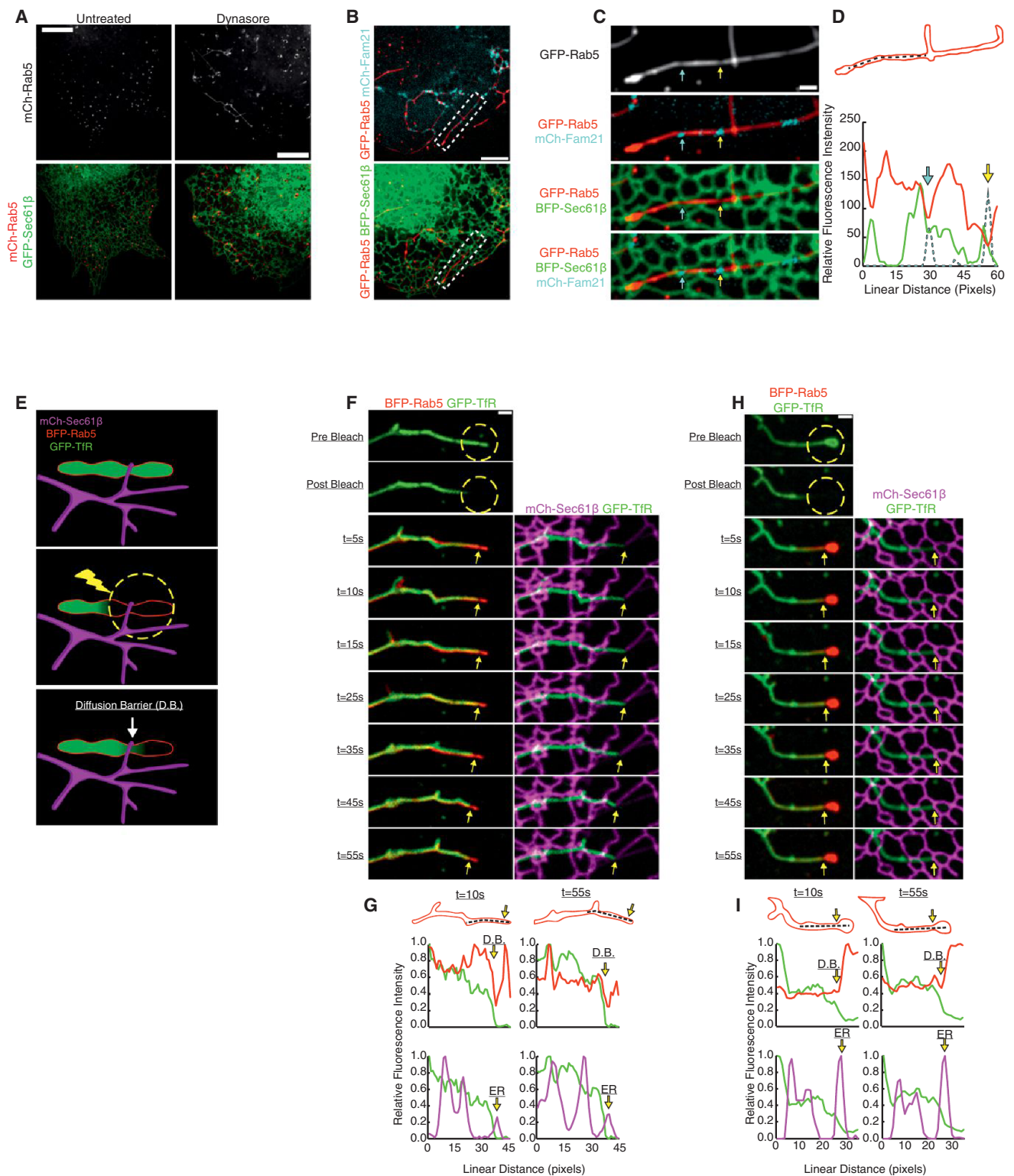


Figure 2. ER Tubules Define Functional Constrictions along Dynasore-Stalled Tubular Endosomes

(A) Images of live Cos-7 cells expressing mCh-Rab5 and GFP-Sec61 β show that dynasore treatment elongates early endosomes while ER is unaffected.

(B) Merged images of a live Cos-7 cell expressing GFP-Rab5, BFP-Sec61 β , and mCh-FAM21 following dynasore treatment.

(C) Magnified images of endosome boxed in (B). Images reveal that FAM21 concentrates at positions along the tubular endosome where Rab5-labeling is reduced and ER tubules intersect these FAM21 domains.

(legend continued on next page)

exists at minimums in fluorescence intensity, which would reflect a functional constriction at ER contact sites (see model Figure 2E). Cells were cotransfected with an ER marker (mCh-Sec61 β), an early endosomal marker (BFP-Rab5), and a membrane-bound cargo marker (GFP-TfR). Next, cells were treated with dynasore to generate elongated endosomes. We photobleached GFP-TfR at one end of a tubular endosome and then visualized fluorescence recovery from the unbleached direction (see model Figure 2E). To confirm tubular endosomes persist throughout this process, we covisualized unbleached BFP-Rab5. Two examples of TfR fluorescence recovery along the tubular endosome are shown in kymographs (Figures 2F and 2H). In the example shown in Figure 2F, the GFP-TfR is photobleached as indicated (in the yellow circle), and the fluorescent signal recovers rapidly (within 10 s) up to a position marked by the yellow arrow. This position corresponds to a contact site with the tip of an ER tubule (see merge with the ER on the right). Furthermore, line-scan analysis demonstrates that the endosome constriction (Rab5) and the diffusion barrier overlap with the ER contact site even as the endosome moves over time (Figure 2G, yellow arrows mark the diffusion barrier [D.B.]; also see Movie S3). In the example shown in Figure 2H, recovery is also rapid up to the position marked by a yellow arrow, which corresponds to a crossing ER tubule. As before, diffusion was limited at the position corresponding to a minimum in Rab5 fluorescence on the endosome, confirming that it is indeed a constriction (Figures 2H and 2I). This suggests that the bulk flow of cargo between compartments is restricted at ER-marked constrictions prior to endosome fission.

Fission of Tubular Early Endosomes Occurs at ER Contact Sites

ER association with tubular endosome constrictions is strikingly reminiscent of contact sites between ER tubules and mitochondrial constrictions that undergo fission (Friedman et al., 2011). We therefore tested if ER-marked constrictions along tubular endosomes are stalled intermediates primed to undergo fission. Cells were cotransfected with mCh-Rab5 and GFP-Sec61 β then treated with dynasore to generate tubular endosomes. To capture fission, dynasore was inactivated with fetal bovine serum (FBS) and cells were imaged live (Figure 3A). As with native early endosomes, the ER contacted elongated endosomes and defined the position of constriction and fission (Figures 3B and 3C). We visualized 31 individual fission events (in nine cells, Figure 3D; Movie S4). ER tubule crossings were present at 80.6% of these fission events (Figure 3D). The mean coverage of the ER network on the tubular endosomes analyzed was 22.36% (Figures 3E and 3F), which demonstrates that ER

contact at the site of fission is not due to chance. Together, these data show that ER contact sites are spatially linked to the position of constriction and fission on early sorting endosomes.

ER Recruitment Is Spatially and Temporally Linked to the Position of Late Endosome Fission

The vast majority of late endosomes (>99%) are also tightly associated with the ER over time (Friedman et al., 2013). We asked whether ER contact sites are also spatially linked to the position of late endosome fission. Cos-7 cells were transfected with markers for late endosomes (mCh-Rab7) and the ER (GFP-Sec61 β) and were imaged live by confocal fluorescence microscopy. Exogenous mCh-Rab7 was expressed at levels similar to endogenous (Figure S2A), and the morphology of labeled late endosomes matched previous reports (Figures 4A and S2B) (Barbero et al., 2002). Cells were also pulse-labeled with EGF (conjugated to Alexa Fluor 647) to mark the internalization and segregation of cargo. We imaged the structure, dynamics, and interactions between the endosomes and the ER over time at the periphery of the cell where both organelles are well resolved. A representative example of a late endosome that undergoes fission is shown in Figures 4B and S2C. Here, a small Rab7⁺ compartment buds off of a larger Rab7⁺ vacuolar compartment. As expected, EGF is retained in the vacuolar compartment during constriction and fission (Figures 4B, and 4C; Movie S5). After merging the time-lapsed images of endosome fission with the location of the tubular ER network, we could observe an ER tubule localized to nearly every fission site (96.5%: $n = 29$ from 24 cells, Figure 4D; see also Figure S2D). In the majority of these cases (93%), the ER was positioned perpendicular to the site of fission. By line scan analysis, it becomes clear that a dynamic ER tubule moves into place to cross over and “cup” the bud just before fission occurs (compare Pre-ER to Pre-Fission scans) (Figure 4B and 4C). These data reveal that ER tubules are both spatially and temporally linked to the position of late endosome fission.

Next, we aimed to generate tubular late endosomes to further resolve the position of ER contact relative to the site of fission. Consistent with previous reports (Derivery et al., 2009), we could generate slightly elongated tubular Rab7-labeled late endosomes by efficiently depleting dynamin-2 (Dnm2) by siRNA treatment (Figure 4H). Cells were also simultaneously cotransfected with mCh-Rab7 and GFP-Sec61 β . Prior to imaging, these cells were pulse-labeled with EGF (conjugated to Alexa Fluor 647) to determine whether elongated endosomes could functionally partition cargo. Live confocal fluorescence microscopy revealed that elongated endosomes in Dnm2-depleted cells were still able to uptake EGF and undergo fission while retaining EGF in the

(D) Line-scan analysis of relative FI of the endosome shown in (C) confirms that FAM21 and ER colocalize with Rab5-labeled endosome constrictions (marked by corresponding blue and yellow arrows).

(E) A model outlines the FRAP technique used to test for a cargo diffusion barrier (D.B.) along tubular endosomes.

(F) A live Cos-7 cell expressing BFP-Rab5, mCh-Sec61 β , and GFP-TfR (membrane-bound cargo) was treated with dynasore to elongate endosomes. Endosomes were photobleached in the region indicated (dotted yellow circle). Images were taken at times indicated during the recovery (see kymograph). See also Movie S3. (G) A graph of relative FI during time points shown reveals a D.B. limits the recovery of TfR at the position of an ER-marked constriction (at yellow arrow). The D.B./constriction maintains contact with the ER over time (marked by yellow arrows).

(H and I) Another example as in (F) and (G). Scale bars represent 5 μ m in (A) and (B) and 1 μ m in (C), (F), and (H).

See also Figure S1.

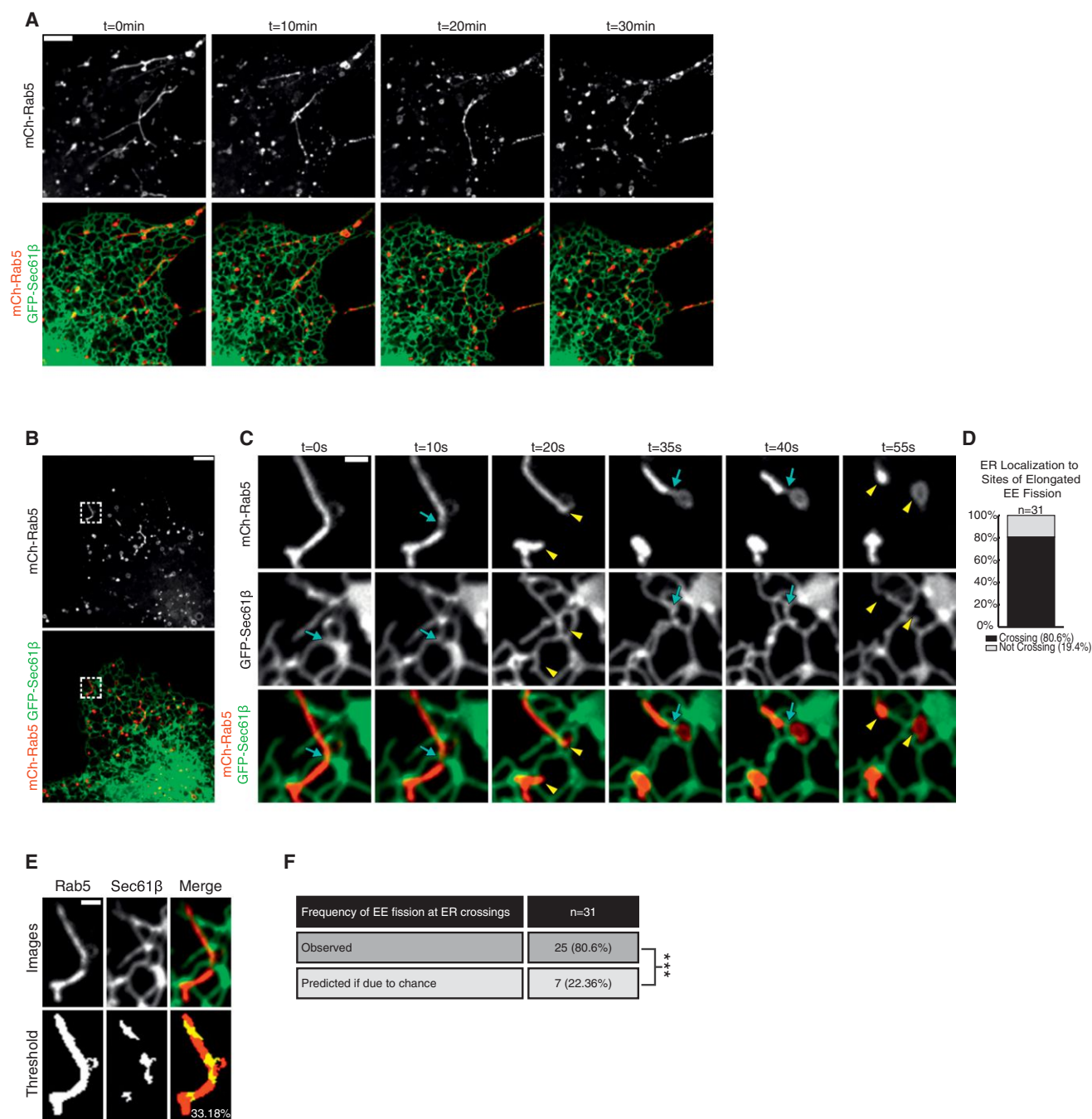


Figure 3. Dynasore-Stalled Tubular Early Endosomes Undergo Fission at ER Contact Sites

(A) The elongated tubular endosome phenotype that results from dynasore treatment can be reversed with the addition of FBS. Time-lapsed images of a live Cos-7 cell expressing mCh-Rab5 and GFP-Sec61 β following dynasore treatment (t = 0 min) and after FBS addition at times indicated (right panels).

(B) Cell treated as in (A).

(C) Magnified time-lapse images of region boxed in B shows two fission events that occur on a single tubular endosome with each event marked by an ER tubule crossing (blue arrows). Yellow arrowheads mark fission products after each division. See also Movie S4.

(D) The percent of tubular EE fission events marked by ER tubule crossings (out of 31 events from nine cells).

(E) Method for determining the amount of endosome image surface covered by ER tubule crossing for all 31 fission events. An example is shown of endosome and ER tracing from an image immediately preceding fission. Top row shows indicated fluorescence markers and bottom row shows thresholded images.

(F) Table shows the percentage (22.36%) of endosomal pixels colocalized with pixels from ER tubules crossing the endosome. The right panel shows that the measured frequency of ER-marked fission (80.6%) is significantly higher than that due to chance (22.36%). ***p < 0.001, Fisher's exact test. Scale bars represent 5 μ m in (A) and (B) and 1 μ m in (C) and (E). min, minutes.

vacuolar compartment (Figures 4E and 4F; Movie S6). In every example recorded (100%, $n = 12$ from eight cells), ER tubules crossed perpendicular to the site of late endosome (LE) fission (Figure 4F). As before, a dynamic ER tubule slides into place just prior to the fission event (compare white and blue arrows in two left panels). We measured the mean coverage of the ER network on tubular endosomes to be 21.24% for the 12 events captured (Figures 4I and 4J). Thus, the average probability of ER-marked fission due to chance (21.24%) is significantly lower than the measured frequency (100%) of ER-marked endosome fission.

ER Tubules Contact Cargo-Sorting Domains Prior to Fission

Next, we aimed to mark endosomal sorting domains prior to fission and ask if this position also coincided with the position of ER recruitment. The retromer complex sorts cargo into a domain on the late endosome that will undergo fission and traffic to the Golgi (Arighi et al., 2004; Seaman, 2004). FAM21 binds and colocalizes with these retromer-mediated sorting domains (Derivery et al., 2009; Gomez and Billadeau, 2009). We thus asked whether ER tubules colocalized with FAM21-marked sorting domains on late endosomes. Cells were transfected with markers of ER (BFP-Sec61 β), late endosomes (GFP-Rab7), and the WASH complex (mCh-FAM21). FAM21 effectively partitions to both punctate and tubular structures on Rab7⁺ late endosomes that are indicative of retromer sorting domains (Figure 5A). We tracked 165 FAM21 structures discretely localized to budding late endosome sorting domains (Figures 5B and 5C; Movie S7). Of these, 80% of FAM21 domains colocalize with the site of ER tubule contact over time (see example in Figure 5B). Thus, most but not all FAM21-marked retromer domains track together with ER tubules, which suggests that the ER is recruited after the sorting domain has formed.

We next observed whether the ER is present at FAM21-marked sorting domains when endosomes are undergoing fission (Figures 5D–5F; Movie S8). The ER was localized to 97% of FAM21-marked fission events ($n = 36$ from 31 cells, Figure 5G). In the example shown, it is important to note that a dynamic ER tubule becomes associated with a FAM21-marked sorting domain just prior to constriction and fission (Figure 5E, compare white and blue arrows at 15 and 20 s, respectively; Movie S8). The temporal recruitment of ER to the position of constriction/fission is also clear by line-scan analysis (Figure 5F, see the peak in ER fluorescence at Pre-Fission). Together, these data demonstrate that ER contact is not required for retromer sorting domain formation or FAM21 recruitment, but that ER contact sites are both spatially and temporally linked to the fission events at FAM21-marked sorting domains.

We have shown that ER tubules make contact with FAM21-marked sorting domains prior to fission. We thus wondered if ER contact with endosomes requires nucleation of actin by the WASH complex. FAM21 binds to the retromer on endosomes; it is the first component of the WASH complex to be recruited, and does so independently of other WASH complex members (Gomez and Billadeau, 2009). We could thus deplete WASH1 which will disrupt WASH complex assembly and actin nucleation (Derivery et al., 2009) and ask whether the ER is still recruited to

FAM21-marked sorting domains and if so whether fission still occurs. We efficiently depleted WASH1 by siRNA and captured 25 FAM21-marked fission events where all but one localized to an ER contact site ($n = 25$ from 21 cells, Figure S3). This demonstrates that the ER is recruited to functional endosome fission sites independent of WASH complex assembly or WASH-mediated actin nucleation.

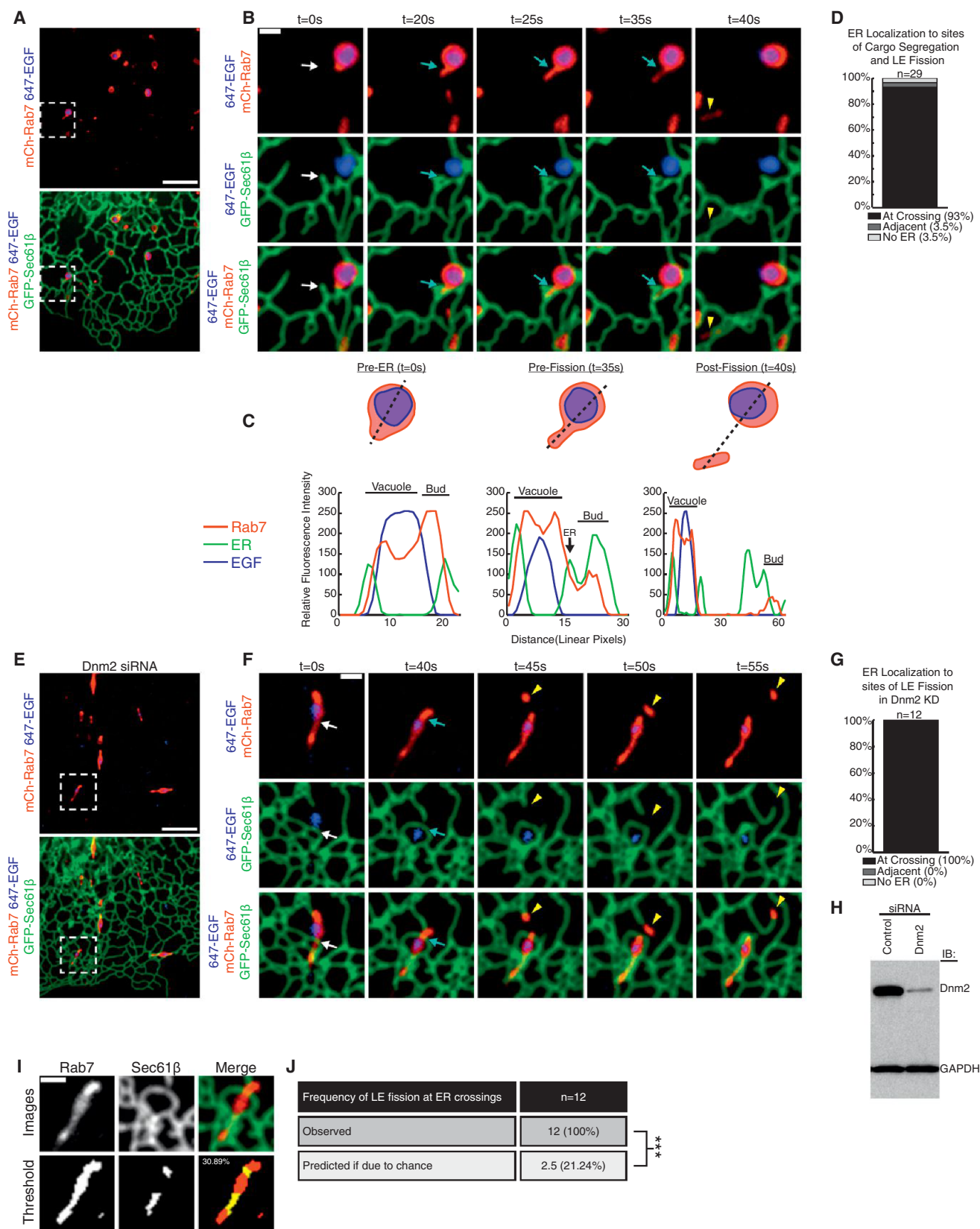
We have shown several examples where a dynamic ER movement establishes contact with a sorting domain immediately prior to fission. ER tubule dynamics are frequent and occur on microtubules (MTs) (Friedman et al., 2010; Waterman-Storer and Salmon, 1998). Thus, we probed the effect of MT depolymerization on ER-endosome contact. Previously, we have shown that endosomes will maintain contact with the ER following MT depolymerization (Friedman et al., 2013). Here, we tested whether ER contacts would also maintain contact with endosome constrictions following MT depolymerization (Figure S4). We show several examples where MTs are depolymerized but ER contact still localizes to endosome constriction sites. We conclude that MT-dependent ER dynamics may be necessary to initiate contact between ER tubules and the constriction site. However, once formed, MTs do not appear necessary to maintain contact.

ER Is Recruited to the Endosome Immediately Prior to Fission

We show that tubular sorting domains can be observed prior to ER contact and fission. Often ER tubules “attack” the site of the preformed sorting domain resulting in fission (Figures 6A and 6B). In the two complimentary examples shown, fission is observed within 10 s of contact formation (Figure 6A and 6B). Many mechanisms have been identified to form the budding domain and sort cargoes (Figure 6C). Microtubule motors have been implicated in sorting and trafficking cargoes (Hunt et al., 2013). Sorting nexins can drive the formation of budding tubules and stabilize them via curvature sensing BAR domains (Cullen, 2008). The retromer complex interacts with the sorting nexins and is involved in sorting of cargoes (Pfeffer, 2009; Seaman et al., 2013). FAM21 binds the retromer and is responsible for recruitment of remaining WASH complex components (Gomez and Billadeau, 2009). The WASH complex stimulates Arp2/3 at endosomal sorting domains thereby activating actin polymerization (Derivery et al., 2009). The presence of actin patches can also help to segregate select cargoes into the bud (Puthenveedu et al., 2010). The role of these multiple factors is to generate sorting domains to traffic cargo. Here, we show that following sorting, the ER establishes contact with the endosome at sites that are spatially and temporally linked to endosome fission.

Disruption of ER Dynamics and Shape Affects the Rate of Endosome Fission

Our data support the following order of events: the budding sorting domain forms, ER tubule contact is established at the interface between the endosome and the bud, and then ~5–10 s later fission occurs at this contact site (Figures 6A and 6B). Consistent with this model, we show several compelling examples where a dynamic ER tubule “attacks” the future fission site. Thus, our



(legend on next page)

prediction was that altering the shape and dynamics of the peripheral ER tubules would lead to a reduction in the frequency of late endosome fission. We chose to disrupt ER shape and dynamics by overexpressing human Reticulon 4a (Rtn4a), which has previously been shown to generate highly elongated and unbranched ER tubules (Shibata et al., 2008; Voeltz et al., 2006). Elongated tubules are coated with immobilized Rtn4a oligomers that can exclude other ER luminal and membrane proteins from the peripheral ER (Shibata et al., 2008; Voeltz et al., 2006; Zurek et al., 2011).

Cells were cotransfected with the general luminal ER marker BFP-KDEL, mCh-Rab7 to visualize endosome fission, and with or without high levels of Rtn4a-GFP. Compared to control, the Rtn4a-GFP cells had an altered ER morphology. As expected, the peripheral ER tubules were highly elongated, unbranched, and luminal BFP-KDEL was mostly excluded from these elongated Rtn4a-GFP tubules (Figure 7A, compare Rtn4a-labeled top panels to control bottom panels). ER tubules were also significantly less dynamic with Rtn4a overexpression. This was measured quantitatively for all cells analyzed using the Pearson's correlation coefficient of overlaid 5 μ m square boxes from two frames taken 60 s apart (Figures 7B and 7C) (French et al., 2008). By this measurement, the time-lapse images of the ER in Rtn4a-GFP expressing cells were more colocalized over time because the peripheral ER moved less, thus resulting in a higher Pearson's correlation (control, $n = 22$; Rtn4a, $n = 20$). We tested whether the change in ER morphology caused by overexpressing Rtn4a impacted endosome fission. We recorded 2 min movies of hundreds of individual late endosomes. For each endosome, we determined whether it formed a budding domain during the course of the movie. For each endosomal bud observed, we measured the size of the endosome, the maximum bud length, and whether it underwent fission (control, $n = 257$; Rtn4a, $n = 257$). We found that altering ER shape did not significantly affect endosome size or bud length (Figures 7D and 7E). However, we found that the number of buds that underwent fission was significantly reduced with Rtn4a overexpression (Figure 7F). Thus, when ER membrane composition, shape, and dynamics are altered, the process of endosome fission is defective.

DISCUSSION

We have shown here that ER tubules form stable contact sites with sorting domains on early and late endosomes and that ER contact is a defining feature of endosome fission. These results are strikingly analogous to previous data showing that ER contact sites define the position of mitochondrial constriction and division (Friedman et al., 2011). Together, these reports demonstrate that the ER has an important and nontraditional role in the cell: to regulate the dynamics and biogenesis of at least two unique cytoplasmic organelles. How could an ER contact site regulate the constriction and/or fission of another organelle? We favor three possibilities. First, contact sites could provide a platform for the recruitment of necessary factors from each organelle that are required to drive fission. Second, because the ER is the site of synthesis for most phospholipids, it is possible that direct translocation of lipids from the ER into the endosomal membrane at contact sites generates a region of high lipid membrane curvature that would promote fission. The ER is also home to lipid modifying enzymes that may transform lipids on the endosomal membrane to similarly drive localized changes in membrane curvature. A third favored model is that ER stores could provide high local concentrations of Ca^{2+} required to activate a Ca^{2+} -dependent activity on the endosome and regulate fission. Indeed, these mechanisms are not mutually exclusive and may all contribute to organelle fission at contact sites.

From an evolutionary point of view, one might think the ability of the ER to divide neighboring organelles would be advantageous for expedited ER dynamics within the cell or for whole-cell rearrangements that occur during mitosis. Both are daunting tasks considering the density of the organelles in the intracellular environment. However, the position of the ER at the site of fission appears to be purposeful for fission of endosomes as well as for mitochondria. We have observed that the ER settles to form stable contact sites with the divide between sorting compartments prior to fission. Likewise, on mitochondria, the ER is also purposefully localized to the position of nucleoid segregation prior to mitochondrial division (Murley et al., 2013).

How ER tubules bind to fission sites on endosomes (and even on mitochondria) is not known and remains an open and

Figure 4. Late Endosome Division Occurs at ER Contact Sites

- (A) A Cos-7 cell expressing mCh-Rab7 (late endosome) and GFP-Sec61 β was pulse-labeled with EGF conjugated to Alexa Fluor 647 (cargo in blue).
 (B) Magnified image of the region boxed in (A) shows an example of late endosome fission. Merged images show the relative location of Rab7, EGF, and the ER, as indicated over time. See also Movie S5.
 (C) Traced outline and the corresponding line-scan analysis of relative FI through the equator of the dividing endosome shown in (B). Relative FI of Rab7, ER, and EGF were performed for time points: $t = 0$ s (Pre-ER), $t = 35$ s (Pre-Fission), and $t = 40$ s (Post-Fission). Note that a dynamic ER tubule is recruited to the position and timing of endosome constriction and fission (compare position of the ER tubule marked by an arrow at $t = 0$ s to $t = 20$ s in B and C).
 (D) Percent of late endosome division events that co-localize with ER tubules ($n = 29$ from 24 cells).
 (E) Image of a Dnm2-depleted cell expressing mCh-Rab7 (red) and GFP-Sec61 β (green) and pulse-labeled with EGF conjugated to Alexa Fluor 647 (blue).
 (F) Magnified image of box in (E) shows an elongated endosome that moves until it becomes associated with an ER tubule (white arrow) at the site of endosome division (blue arrow). The location of the Rab7 endosome bud (yellow arrowhead) is shown. See also Movie S6.
 (G) The percent of late endosome fission events that colocalize with ER tubules in the absence of dynamin-2 (out of 12 events from eight cells).
 (H) Immunoblot analysis shows efficient depletion of Dnm2 in cells transfected with Dnm2 siRNA (right lane) relative to control (left lane).
 (I) The endosome image surface covered by ER tubule crossing was measured for all 12 fission events. In example shown, the top row shows indicated fluorescence markers and bottom row shows thresholded images.
 (J) Table summarizes predicted frequency of ER-marked LE fission (21.24% based on coverage) versus the actual frequency of ER-marked fission (100%).
 *** $p < 0.001$, Fisher's exact test. Scale bars represent 5 μ m in (A) and (E) and 1 μ m in (B) and (F).
 See also Figure S2.

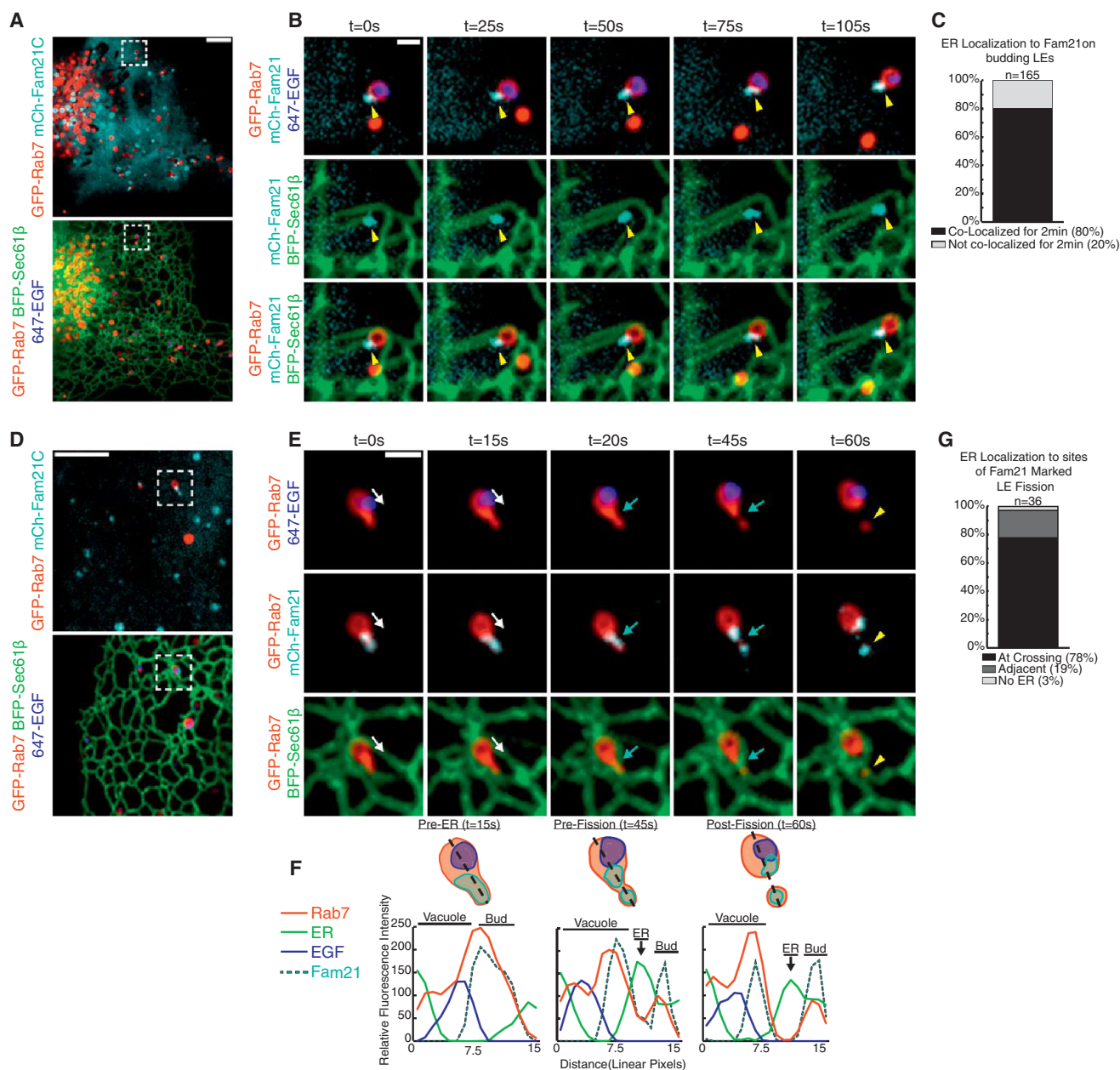


Figure 5. The ER Is Recruited to FAM21-Marked Sorting Domains Prior to Fission

(A) Merged image of a live Cos-7 cell expressing GFP-Rab7, BFP-Sec61 β , and mCh-FAM21 that was also pulse-labeled with EGF conjugated to Alexa Fluor 647 (cargo in blue).

(B) Magnified time-lapse images of an endosome in (A) shows the tip of an ER tubule tracking with a FAM21 punctum at the bud of a late endosome (yellow arrowhead). See also Movie S7.

(C) The number of FAM21 puncta on late endosome buds that maintain contact with ER over a 2 min time course (165 puncta from 31 cells).

(D) Merged image of a live Cos-7 treated as in (A).

(E) Magnified images of boxed region in (D) shows a late endosome undergoing fission. A bud labeled by FAM21 extends from the late endosome and undergoes constriction and division. An ER tubule (follow white arrow) is recruited to the FAM21 bud just before fission (at the blue arrow). The location of the exiting Rab7 endosome bud is marked by a yellow arrowhead. See also Movie S8.

(F) Line-scan analysis of endosome shown in E relates the timing and position of ER tubule recruitment relative to Rab7, FAM21, and EGF.

(G) The percent of FAM21-marked late endosome division events that colocalize with ER tubules during cargo segregation, (n = 36 from 31 cells). Scale bars represent 5 μ m in (A) and (D) and 1 μ m in (B) and (E).

See also Figures S3 and S4.

important question to answer. Although several partners have been identified that can bridge the apposing membranes, including PTP1B-EGFR, VAP-A-ORP1L, and STARD3-ORP1L (Alpy et al., 2013; Eden et al., 2010; Rocha et al., 2009), none have been shown to be required for maintenance of contact. Thus, these could all be independent tethers for independent functions or these complexes could function at sites that are tethered by yet unidentified factors. There are likely multiple points of contact between the ER and endosomes that could coordinate functions. For example, previous immunoelectron microscopy has also shown colocalization of EGFR and PTP1B at the ER-endosome interface (Eden et al., 2010; Haj et al., 2002). By live confocal microscopy, we also find that EGF within endosomes frequently tracks with the position of ER contact (Figure S5A). Therefore, we tested if ER-marked endosome fission occurs at the same contact sites as EGF clustering. We found that EGF puncta localized to less than half of the ER-marked fission events (Figure S5B). These data demonstrate that multiple functional and structural contact sites exist between the ER and endosomes to coordinate different functions during cargo sorting and fission. The machinery at ER-marked endosome fission sites could be especially difficult to identify because, in many examples, these contacts are established immediately prior to the fission event and thus are likely to be very transient. The identification of this transient machinery will be a next major goal.

EXPERIMENTAL PROCEDURES

Plasmid Construction

GFP-Sec61 β (Shibata et al., 2008), BFP-Sec61 β (Zurek et al., 2011), mCh-Sec61 β (Zurek et al., 2011), mCh-Rab5B (Friedman et al., 2010), BFP-Rab5B (Friedman et al., 2013), Rtn4a-GFP (Shibata et al., 2008), and mCh- α Tubulin (Friedman et al., 2010) were previously described. GFP-Rab4B was a gift from A.R. English (University of Colorado-Boulder) and was generated by PCR amplifying human Rab4B (National Center for Biotechnology Information [NCBI] accession number NM_016154.4) into BglII/KpnI of pAcGFP1-C1 (Clontech). GFP-Rab5B was a gift from J.R. Friedman (University of California-Davis) and was generated by PCR amplifying human Rab5B (NCBI accession number NM_002868.3) and cloning it into the XhoI/BamHI sites of pAcGFP1-C1 (Clontech). To generate GFP-Rab7A, human Rab7A (NCBI accession number NM_004637.5) was PCR amplified and cloned into XhoI/HindIII sites of the pAcGFP1-C1 vector. mCh-Rab7A was made like GFP-Rab7A but mCh was cloned into the NheI/XhoI sites of pAcGFP1-C1, replacing the GFP. mCh-FAM21 and BFP-FAM21 were generated by subcloning mCh- or BFP- into the BamHI and MluI sites while removing HA-YFP from the shFAM21/HA-YFP-FAM21 rescue vector given to us by Dr. Daniel Billadeau described in Gomez and Billadeau (2009). GFP-TfR was generated by Dr. Lois Greene.

Cell Growth and Plasmid DNA Transfections

Cos-7 cells (ATCC) were grown in DMEM media (Invitrogen) supplemented with 10% FBS and 1% penicillin/streptomycin. Cells were seeded in a 60 \times 15 mm dish at 5.0×10^5 cells \sim 16 hr prior to transfection. Cells were transfected \sim 24 hr prior to imaging with plasmid DNA in Opti-MEM media (Invitrogen) with 12.5 μ l of Lipofectamine 2000 reagent (Invitrogen) according to the manufacturer's instructions. After \sim 5 hr of transfection, cells were seeded in 35 mm glass-bottom microscope dishes (MatTek) at 2.0×10^5 cells. Cells were imaged in 37°C Opti-MEM. For all experiments, the following amounts of plasmid DNA were transfected into cells for experiments: 500 ng/ml BFP-Sec61 β ; 250 ng/ml mCh-Sec61 β and GFP-Sec61 β ; 25 ng/ml GFP-Rab4; 50 ng/ml BFP-Rab5 and GFP-Rab5; 125 ng/ml mCh-Rab5; 20 ng/ml

mCh-Rab7; 250 ng/ml mCh-FAM21; 500 ng/ml BFP-FAM21; 165 ng/ml GFP-TfR; 500 ng/ml Rtn4a-GFP; 250 ng/ml BFP-KDEL; and 62.5 ng/ml mCh- α Tubulin.

RNAi Transfection and Western Blot

Dynamin-2 was depleted using an ON-TARGETplus Human DNM2 (1785) siRNA SMARTpool (Thermo). WASH was depleted using ON-TARGETplus Human WASH1 (100287171) siRNA SMARTpool (Thermo). Cells were seeded in a 60 \times 15 mm dish at 5.0×10^5 cells \sim 16 hr prior to first transfection. Cells were first transfected \sim 72 hr prior to imaging with 12.5 μ l Dharmatect (Thermo) in DMEM with 10% FBS and 25 nM RNAi oligonucleotides or 25 nM Silencer Negative Control #1 siRNA (Ambion AM4635). After \sim 6 hr of transfection, cells were washed and media was replaced with DMEM supplemented with 10% FBS and 1% penicillin/streptomycin. Cells were transfected again \sim 24 hr prior to imaging with plasmid DNA as described before with the addition of 25 nM RNAi oligonucleotides or 25 nM Silencer Negative Control #1 siRNA. After \sim 5 hr of transfection, cells were seeded in 35 mm glass-bottom microscope dishes (MatTek) at 2.0×10^5 cells. Cells were imaged in 37°C Opti-MEM.

Whole cell lysates of Cos-7 cells were resuspended in Laemmli sample buffer, boiled for 10 min, separated by SDS-PAGE, and transferred to a PVDF membrane. Primary antibody concentrations were used as follows: Rab7 (Cell Signaling 99367S) 1:1,000; Dynamin-2 (Abcam ab3457) 1:15,000; WASH1 (gift from D. Billadeau) 1:6,000; GAPDH (Sigma-Aldrich G9545) was used as a loading control at 1:20,000. HRP-conjugated goat anti-rabbit secondary antibody (Sigma-Aldrich) was used at 1:3,000, and signal was detected with SuperSignal West Pico Chemiluminescent Substrate (Thermo). Band density was estimated using ImageJ (NIH) Gel Analyzer.

Confocal Microscopy

Live-cell imaging was performed with an inverted fluorescence microscope (TE2000-U; Nikon) equipped with an electron-multiplying charge-coupled device (CCD) camera (Cascade II; Photometrics) and a Yokogawa spinning-disk confocal system (CSU-Xm2; Nikon). Images were taken with a 100 \times numerical aperture 1.4 oil objective (Nikon). While imaging, live cells were kept at 37°C in a live-cell incubation chamber (Pathology Devices). Images were acquired with MetaMorph 7.0 (MDS Analytical Technologies), analyzed, merged, and contrasted using ImageJ, as well as contrasted and converted to 400 dpi using Photoshop (Adobe). Scale bars were generated using ImageJ. Supplemental videos were generated using ImageJ.

Endosome Fission and ER Coverage Analysis

Cos-7 cells expressing markers for ER and endosomes were visualized live by confocal fluorescence microscopy every 5 s for 2 min in a single focal plane. Fission events were counted if they were in regions of the cell with resolvable ER tubules. Association of ER tubules was classified as adjacent, crossing, or not associated. Line-scans were performed using ImageJ by drawing a line the width of the budding compartment along the length of the entire endosome. Fission of dynasore-induced elongated endosomes (Figures 3B–3D) and Dynamin-2 siRNA-induced elongated endosomes (Figures 4E–4G) was analyzed if ER coverage was measured as $<35\%$ of the total endosome area.

To predict the frequency of ER at the site of fission if fission could occur at any ER crossing along the endosome, we measured the area of crossing ER tubules versus total endosome area from the image immediately preceding endosome fission for all events. The area was measured using the adaptive threshold plugin for ImageJ (created by Qingzong Tseng) which converts the 8-bit images to binary form defining the ER or endosome structure. We found the percent of endosome area covered by crossing ER tubules by comparing the number of pixels positive for ER signal to total number of pixels from the endosome area. We only selected the regions of ER that crossed the elongated endosome, omitting any ER tubules that ran parallel to the endosome.

Fluorescent Cargo Treatment and Imaging

Prior to imaging, cells were incubated in 37°C growth media containing either 1 μ g/ml EGF conjugated to Alexa Fluor 647 (Invitrogen) or 50 μ g/ml transferrin conjugated to Alexa Fluor 647 (Invitrogen) for 5–10 min. Then, cells were

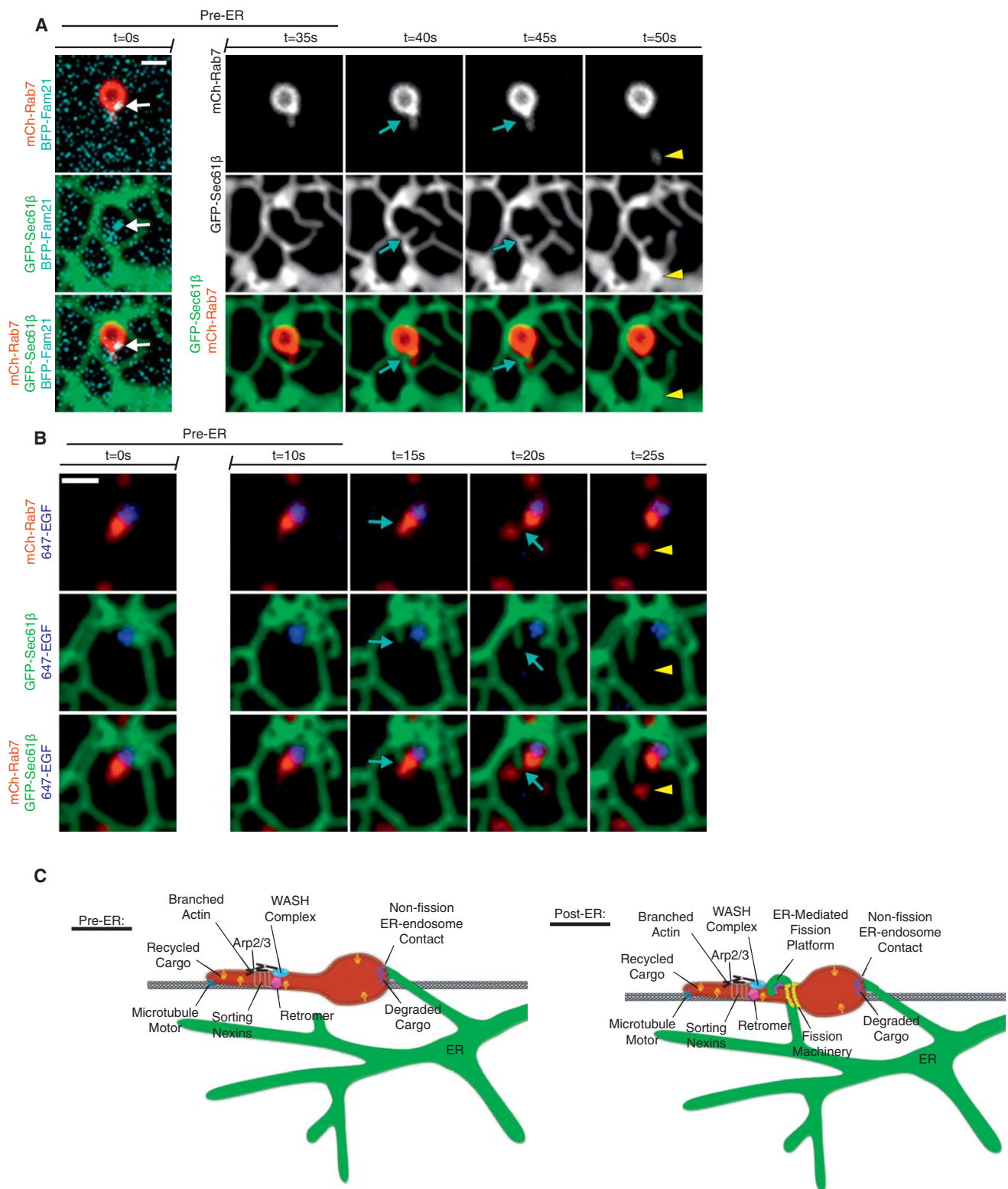


Figure 6. Model of Endosome Sorting and Fission Factors

(A) A live Cos-7 cell expressing mCh-Rab7, GFP-Sec61 β , and BFP-FAM21. Endosome fission occurs after a dynamic ER tubule contacts the endosome at a constricted sorting domain.

(legend continued on next page)

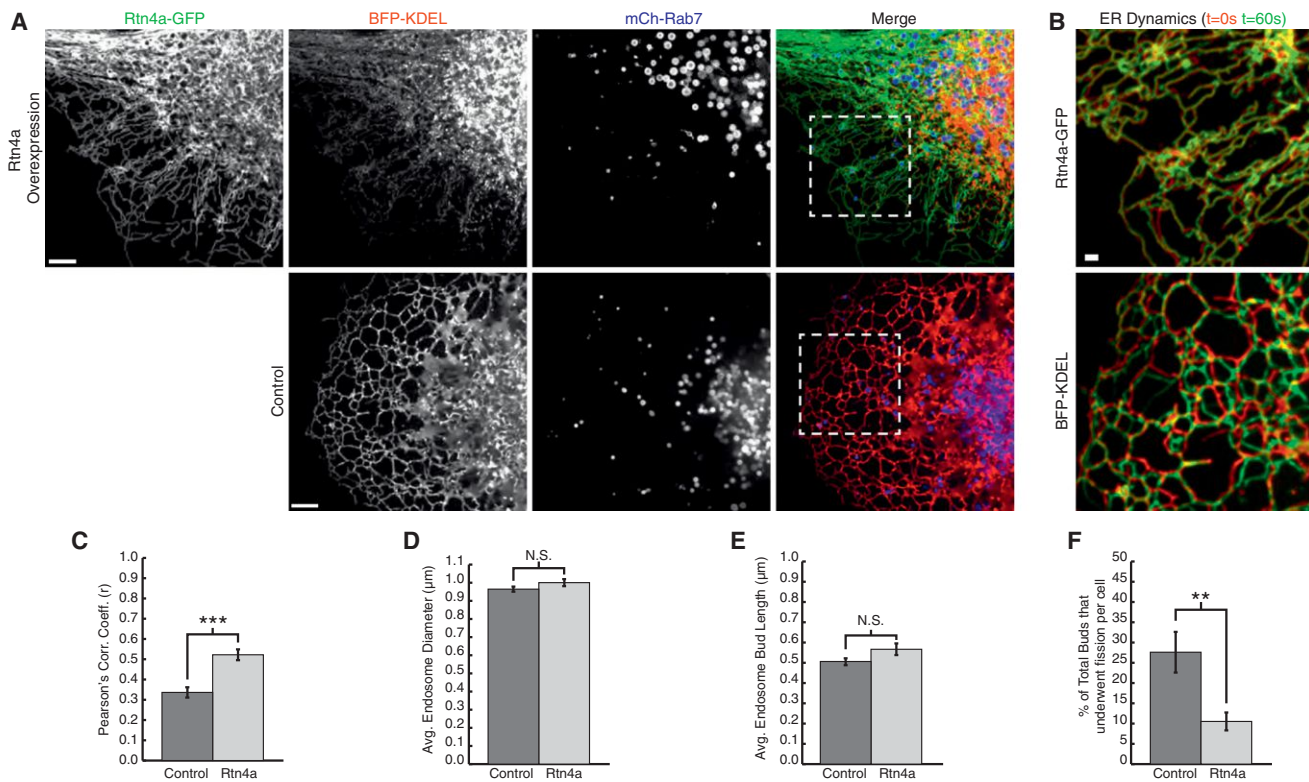


Figure 7. ER Structure and Dynamics Affect Endosome Fission

(A) Live Cos-7 cells expressing BFP-KDEL and mCh-Rab7 with (top panel) or without (bottom panel) Rtn4a-GFP. Note that Rtn4a-GFP tubules are less branched (top first panel) and exclude KDEL from the periphery (top second panel).
 (B) Zoom from (A) shows overlay of $t = 0$ s (red) with $t = 60$ s (green) to show ER movement over time “dynamics.”
 (C) Graph of average Pearson’s correlation coefficient shows greater colocalization and therefore fewer dynamic ER movements for Rtn4a expressing cells (control, $n = 22$ cells; Rtn4a, $n = 20$ cells; *** $p < 0.001$, two-tailed t test).
 (D) Graph showing no significant difference in endosome diameter between control and Rtn4a expressing cells (control, $n = 216$ endosomes; Rtn4a, $n = 215$ endosomes).
 (E) Graph showing no significant difference in endosome bud length between control and Rtn4a expressing cells (control, $n = 257$ buds in 22 cells; Rtn4a, $n = 257$ buds in 20 cells).
 (F) Graph showing a significant (** $p < 0.01$, two-tailed t test) decrease in the percent of endosome buds from (E) that undergo productive fission in Rtn4a expressing cells. Scale bars represent $5 \mu\text{m}$ in (A) and $1 \mu\text{m}$ (B) and (G). Error bars represent SEM.

washed two times in 37°C Opti-MEM and imaged for less than 30 min to localize transferrin or EGF to early endosomes or after 30 min to localize EGF to late endosomes.

Dynasore Treatment and Imaging

Cells were passaged and transfected as described above. Cells were then washed two times and incubated in serum-free, antibiotic-free DMEM containing $80\text{--}160 \mu\text{M}$ dynasore (Sigma-Aldrich) for 30 min prior to imaging. After locating cells containing tubular endosomes by confocal microscopy, 10% FBS was added to inactivate dynasore thereby preventing adverse effects and restoring fission.

To label cargo in dynasore-treated cells, cells were simultaneously incubated with either $1 \mu\text{g/ml}$ EGF conjugated to Alexa Fluor 647 (Invitrogen) or $50 \mu\text{g/ml}$ transferrin conjugated to Alexa Fluor 647 (Invitrogen) for the duration of the dynasore incubation saturating the endosomes with cargo. Following

the incubation, cells were washed two times with serum-free, antibiotic-free DMEM to remove excess cargo and the media was replaced with serum-free, antibiotic-free DMEM containing $80\text{--}160 \mu\text{M}$ dynasore for imaging.

Diffusion Barrier Identification by Photobleaching

Cos-7 cells were transfected with GFP-TfR, BFP-Rab5, and mCh-Sec61 β as described above. Cells were dynasore-treated as described above. Once a cell containing tubular endosomes was identified, a circular region of interest 30 pixels in diameter was created using MetaMorph 7.0. Using the Mosaic Digital Diaphragm System for Nikon TE2000 Microscope (Photonic Instruments), an image was captured prebleach and the region of interest was photobleached for 2 s (Uniblitz VCM-D1 shutter, Vincent Associates) using an Argon laser with a wavelength of $450\text{--}515 \text{ nm}$ (National Laser). An image was taken postbleach followed by a 1 min movie of all expressed channels with 5 s intervals.

(B) A Cos-7 cell expressing mCh-Rab7 and GFP-Sec61 β was pulse-labeled with EGF conjugated to Alexa Fluor 647. The ER tubule contacts the endosome, fission occurs at the point of contact, the budding domain leaves, maintaining ER contact.

(C) The many factors implicated in endosome sorting and fission are indicated in the model pre- and post-ER recruitment. Following sorting, the ER establishes contact with the endosome at sites that are spatially and temporally linked to endosome fission.

See also Figure S5.

To analyze the pattern of recovery and thus any barriers to diffusion, line-scans and kymographs were created using ImageJ. We identified constrictions marked by ER tubule crossings by drawing a line the width of the tubular endosome being analyzed and measuring the mean fluorescence intensity along the length of the tubular endosome.

Measuring Dynamic ER Movements

Live Cos-7 cells were imaged every 5 s for 2 min. Raw images taken 60 s apart were converted to 8 bit and overlaid using ImageJ. The average Pearson's correlation coefficient was determined for three peripheral 5 μ m boxes for each cell using the PSC Colocalization plugin for ImageJ (French et al., 2008). A higher degree of colocalization (Pearson's correlation coefficient closer to 1.0) between the images over time is indicative of a less dynamic ER.

Nocodazole Treatment and Imaging

Dynamin-2 was depleted as described above to enrich for elongated endosomes. Cells were passaged and transfected as described above. Cells were then washed and incubated in 37°C Opti-MEM prior to imaging. After locating cells containing elongated endosomes by confocal microscopy, a z stack was acquired. Then, 5 μ M nocodazole (Acros Organics) was added for at least 15 min or until the majority of microtubules had depolymerized. A z stack was taken posttreatment to observe the position of the ER relative to endosome constrictions.

SUPPLEMENTAL INFORMATION

Supplemental Information includes five figures and eight movies and can be found with this article online at <http://dx.doi.org/10.1016/j.cell.2014.10.023>.

AUTHOR CONTRIBUTIONS

A.A.R., P.J.C., and G.K.V. contributed to experimental design. A.A.R. and P.J.C. conducted all experiments, data analysis, and figure composition. M.J.P. generated expression constructs and assisted with experiments. A.A.R. and G.K.V. wrote the manuscript. A.A.R., P.J.C., M.J.P., and G.K.V. proofed the manuscript.

ACKNOWLEDGMENTS

We thank J. Nunnari, J. Friedman, and L. Greene for helpful discussions. We also thank C. Merrifield, L. Greene, and G. Hajnoczky for generating plasmids used here and D. Billadeau for the FAM21 containing plasmid and the WASH1 antibody. A.A.R. is a Gordon Stone Fellow and was supported by a pre-doctoral training grant from the NIH (GM07135). M.J.P. is supported by an NSF Graduate Research Fellowship DGE 1144083 and was supported by a pre-doctoral training grant from the NIH (T32 GM08759). This work is also supported by a Research Scholar grant from the American Cancer Society and by a grant from the NIH (GM083977) to G.K.V.

Received: May 16, 2014

Revised: August 13, 2014

Accepted: September 24, 2014

Published: November 20, 2014

REFERENCES

- Alpy, F., Rousseau, A., Schwab, Y., Legueux, F., Stoll, I., Wendling, C., Spiegelhalter, C., Kessler, P., Mathelin, C., Rio, M.-C., et al. (2013). STARD3 or STARD3NL and VAP form a novel molecular tether between late endosomes and the ER. *J. Cell Sci.* 126, 5500–5512.
- Arighi, C.N., Hartnell, L.M., Aguilar, R.C., Haft, C.R., and Bonifacio, J.S. (2004). Role of the mammalian retromer in sorting of the cation-independent mannose 6-phosphate receptor. *J. Cell Biol.* 165, 123–133.
- Barbero, P., Bittova, L., and Pfeffer, S.R. (2002). Visualization of Rab9-mediated vesicle transport from endosomes to the trans-Golgi in living cells. *J. Cell Biol.* 156, 511–518.
- Cullen, P.J. (2008). Endosomal sorting and signalling: an emerging role for sorting nexins. *Nat. Rev. Mol. Cell Biol.* 9, 574–582.
- Derivery, E., Sousa, C., Gautier, J.J., Lombard, B., Loew, D., and Gautreau, A. (2009). The Arp2/3 activator WASH controls the fission of endosomes through a large multiprotein complex. *Dev. Cell* 17, 712–723.
- Dunn, K.W., McGraw, T.E., and Maxfield, F.R. (1989). Iterative fractionation of recycling receptors from lysosomally destined ligands in an early sorting endosome. *J. Cell Biol.* 109, 3303–3314.
- Eden, E.R., White, I.J., Tsapara, A., and Futter, C.E. (2010). Membrane contacts between endosomes and ER provide sites for PTP1B-epidermal growth factor receptor interaction. *Nat. Cell Biol.* 12, 267–272.
- Elbaz, Y., and Schuldiner, M. (2011). Staying in touch: the molecular era of organelle contact sites. *Trends Biochem. Sci.* 36, 616–623.
- French, A.P., Mills, S., Swarup, R., Bennett, M.J., and Pridmore, T.P. (2008). Colocalization of fluorescent markers in confocal microscope images of plant cells. *Nat. Protoc.* 3, 619–628.
- Friedman, J.R., Webster, B.M., Mastronarde, D.N., Verhey, K.J., and Voeltz, G.K. (2010). ER sliding dynamics and ER-mitochondrial contacts occur on acetylated microtubules. *J. Cell Biol.* 190, 363–375.
- Friedman, J.R., Lackner, L.L., West, M., DiBenedetto, J.R., Nunnari, J., and Voeltz, G.K. (2011). ER tubules mark sites of mitochondrial division. *Science* 334, 358–362.
- Friedman, J.R., DiBenedetto, J.R., West, M., Rowland, A.A., and Voeltz, G.K. (2013). Endoplasmic reticulum-endosome contact increases as endosomes traffic and mature. *Mol. Biol. Cell* 24, 1030–1040.
- Gautreau, A., Oguievetskaia, K., and Ungermann, C. (2014). Function and regulation of the endosomal fusion and fission machineries. *Cold Spring Harb. Perspect. Biol.* 6, 1–16.
- Gomez, T.S., and Billadeau, D.D. (2009). A FAM21-containing WASH complex regulates retromer-dependent sorting. *Dev. Cell* 17, 699–711.
- Haj, F.G., Verveer, P.J., Squire, A., Neel, B.G., and Bastiaens, P.I.H. (2002). Imaging sites of receptor dephosphorylation by PTP1B on the surface of the endoplasmic reticulum. *Science* 295, 1708–1711.
- Hanyaloglu, A.C., and von Zastrow, M. (2008). Regulation of GPCRs by endocytic membrane trafficking and its potential implications. *Annu. Rev. Pharmacol. Toxicol.* 48, 537–568.
- Harbour, M.E., Breusegem, S.Y., and Seaman, M.N.J. (2012). Recruitment of the endosomal WASH complex is mediated by the extended 'tail' of Fam21 binding to the retromer protein Vps35. *Biochem. J.* 442, 209–220.
- Helle, S.C.J., Kanfer, G., Kolar, K., Lang, A., Michel, A.H., and Kornmann, B. (2013). Organization and function of membrane contact sites. *Biochim. Biophys. Acta* 1833, 2526–2541.
- Hunt, S.D., Townley, A.K., Danson, C.M., Cullen, P.J., and Stephens, D.J. (2013). Microtubule motors mediate endosomal sorting by maintaining functional domain organization. *J. Cell Sci.* 126, 2493–2501.
- Jia, D., Gomez, T.S., Metlagel, Z., Umetani, J., Otwinowski, Z., Rosen, M.K., and Billadeau, D.D. (2010). WASH and WAVE actin regulators of the Wiskott-Aldrich syndrome protein (WASP) family are controlled by analogous structurally related complexes. *Proc. Natl. Acad. Sci. USA* 107, 10442–10447.
- Kilpatrick, B.S., Eden, E.R., Schapira, A.H., Futter, C.E., and Patel, S. (2013). Direct mobilisation of lysosomal Ca²⁺ triggers complex Ca²⁺ signals. *J. Cell Sci.* 126, 60–66.
- Kirchhausen, T., Macia, E., and Pelish, H.E. (2008). Use of dynasore, the small molecule inhibitor of dynamin, in the regulation of endocytosis. *Methods Enzymol.* 438, 77–93.
- López-Sanjurjo, C.I., Tovey, S.C., Prole, D.L., and Taylor, C.W. (2013). Lysosomes shape Ins(1,4,5)P₃-evoked Ca²⁺ signals by selectively sequestering Ca²⁺ released from the endoplasmic reticulum. *J. Cell Sci.* 126, 289–300.
- Macia, E., Ehrlich, M., Massol, R., Boucrot, E., Brunner, C., and Kirchhausen, T. (2006). Dynasore, a cell-permeable inhibitor of dynamin. *Dev. Cell* 10, 839–850.

- Maxfield, F.R., and McGraw, T.E. (2004). Endocytic recycling. *Nat. Rev. Mol. Cell Biol.* 5, 121–132.
- Mayor, S., Presley, J.F., and Maxfield, F.R. (1993). Sorting of membrane components from endosomes and subsequent recycling to the cell surface occurs by a bulk flow process. *J. Cell Biol.* 121, 1257–1269.
- Mesaki, K., Tanabe, K., Obayashi, M., Oe, N., and Takei, K. (2011). Fission of tubular endosomes triggers endosomal acidification and movement. *PLoS ONE* 6, e19764.
- Morgan, A.J., Davis, L.C., Wagner, S.K.T.Y., Lewis, A.M., Parrington, J., Churchill, G.C., and Galione, A. (2013). Bidirectional Ca^{2+} signaling occurs between the endoplasmic reticulum and acidic organelles. *J. Cell Biol.* 200, 789–805.
- Murley, A., Lackner, L.L., Osman, C., West, M., Voeltz, G.K., Walter, P., and Nunnari, J. (2013). ER-associated mitochondrial division links the distribution of mitochondria and mitochondrial DNA in yeast. *eLife* 2, e00422.
- Park, R.J., Shen, H., Liu, L., Liu, X., Ferguson, S.M., and De Camilli, P. (2013). Dynamin triple knockout cells reveal off target effects of commonly used dynamin inhibitors. *J. Cell Sci.* 126, 5305–5312.
- Pfeffer, S.R. (2009). Multiple routes of protein transport from endosomes to the trans Golgi network. *FEBS Lett.* 583, 3811–3816.
- Puthenveedu, M.A., Lauffer, B., Temkin, P., Vistein, R., Carlton, P., Thorn, K., Taunton, J., Weiner, O.D., Parton, R.G., and von Zastrow, M. (2010). Sequence-dependent sorting of recycling proteins by actin-stabilized endosomal microdomains. *Cell* 143, 761–773.
- Rocha, N., Kuijl, C., van der Kant, R., Janssen, L., Houben, D., Janssen, H., Zwart, W., and Neefjes, J. (2009). Cholesterol sensor ORP1L contacts the ER protein VAP to control Rab7-RILP-p150 Glued and late endosome positioning. *J. Cell Biol.* 185, 1209–1225.
- Seaman, M.N.J. (2004). Cargo-selective endosomal sorting for retrieval to the Golgi requires retromer. *J. Cell Biol.* 165, 111–122.
- Seaman, M.N.J. (2008). Endosome protein sorting: motifs and machinery. *Cell. Mol. Life Sci.* 65, 2842–2858.
- Seaman, M.N.J., Gautreau, A., and Billadeau, D.D. (2013). Retromer-mediated endosomal protein sorting: all WASHed up!. *Trends Cell Biol.* 23, 522–528.
- Shibata, Y., Voss, C., Rist, J.M., Hu, J., Rapoport, T.A., Prinz, W.A., and Voeltz, G.K. (2008). The reticulon and DP1/Yop1p proteins form immobile oligomers in the tubular endoplasmic reticulum. *J. Biol. Chem.* 283, 18892–18904.
- Sönnichsen, B., De Renzis, S., Nielsen, E., Rietdorf, J., and Zerial, M. (2000). Distinct membrane domains on endosomes in the recycling pathway visualized by multicolor imaging of Rab4, Rab5, and Rab11. *J. Cell Biol.* 149, 901–914.
- Stefan, C.J., Manford, A.G., and Emr, S.D. (2013). ER-PM connections: sites of information transfer and inter-organelle communication. *Curr. Opin. Cell Biol.* 25, 434–442.
- Toulmay, A., and Prinz, W.A. (2011). Lipid transfer and signaling at organelle contact sites: the tip of the iceberg. *Curr. Opin. Cell Biol.* 23, 458–463.
- Trischler, M., Stoorvogel, W., and Ullrich, O. (1999). Biochemical analysis of distinct Rab5- and Rab11-positive endosomes along the transferrin pathway. *J. Cell Sci.* 112 (Pt 24), 4773–4783.
- van der Kant, R., and Neefjes, J. (2014). Small regulators, major consequences - Ca^{2+} and cholesterol at the endosome-ER interface. *J. Cell Sci.* 127, 929–938.
- Voeltz, G.K., Prinz, W.A., Shibata, Y., Rist, J.M., and Rapoport, T.A. (2006). A class of membrane proteins shaping the tubular endoplasmic reticulum. *Cell* 124, 573–586.
- Waterman-Storer, C.M., and Salmon, E.D. (1998). Endoplasmic reticulum membrane tubules are distributed by microtubules in living cells using three distinct mechanisms. *Curr. Biol.* 8, 798–806.
- Zajac, A.L., Goldman, Y.E., Holzbaur, E.L.F., and Ostap, E.M. (2013). Local cytoskeletal and organelle interactions impact molecular-motor- driven early endosomal trafficking. *Curr. Biol.* 23, 1173–1180.
- Zurek, N., Sparks, L., and Voeltz, G. (2011). Reticulon short hairpin transmembrane domains are used to shape ER tubules. *Traffic* 12, 28–41.

The Structural Basis of Substrate Recognition by the Eukaryotic Chaperonin TRiC/CCT

Lukasz A. Joachimiak,^{1,6,*} Thomas Walzthoeni,^{3,4} Corey W. Liu,² Ruedi Aebersold,^{3,5} and Judith Frydman^{1,*}

¹Department of Biology and Genetics, Stanford University, Stanford, CA 94305, USA

²Stanford Magnetic Resonance Laboratory, Stanford University, Stanford, CA 94305, USA

³Institute of Molecular Systems Biology, Department of Biology, ETH Zurich, 8093 Zurich, Switzerland

⁴Ph.D. Program in Molecular Life Sciences, University of Zurich/ETH Zurich, 8057 Zurich, Switzerland

⁵Faculty of Science, University of Zurich, 8006 Zurich, Switzerland

⁶Present address: Pacific Biosciences, 1380 Willow Street, Menlo Park, CA 94025, USA

*Correspondence: lajoachi@gmail.com (L.A.J.), jfrydman@stanford.edu (J.F.)

<http://dx.doi.org/10.1016/j.cell.2014.10.042>

SUMMARY

The eukaryotic chaperonin TRiC (also called CCT) is the obligate chaperone for many essential proteins. TRiC is hetero-oligomeric, comprising two stacked rings of eight different subunits each. Subunit diversification from simpler archaeal chaperonins appears linked to proteome expansion. Here, we integrate structural, biophysical, and modeling approaches to identify the hitherto unknown substrate-binding site in TRiC and uncover the basis of substrate recognition. NMR and modeling provided a structural model of a chaperonin-substrate complex. Mutagenesis and crosslinking-mass spectrometry validated the identified substrate-binding interface and demonstrate that TRiC contacts full-length substrates combinatorially in a subunit-specific manner. The binding site of each subunit has a distinct, evolutionarily conserved pattern of polar and hydrophobic residues specifying recognition of discrete substrate motifs. The combinatorial recognition of polypeptides broadens the specificity of TRiC and may direct the topology of bound polypeptides along a productive folding trajectory, contributing to TRiC's unique ability to fold obligate substrates.

INTRODUCTION

The health and integrity of the cellular proteome depend on molecular chaperones, which through their distinct substrate specificities and modes of action maintain protein homeostasis (Balch et al., 2008; Kim et al., 2013; Li and Buchner, 2013; Saibil, 2013). Among these, the eukaryotic chaperonin TRiC (for TCP-1 ring complex, also called CCT for chaperonin containing TCP1) is distinguished by its complex architecture and mechanism, which allow it to fold a subset of essential and topologically complex proteins, including cell-cycle regulators, signaling proteins, and cytoskeletal components (Bigotti and Clarke, 2008; Kim et al., 2013).

TRiC/CCT is a large hetero-oligomeric ATP-dependent complex consisting of two eight-membered rings stacked back to back (Bigotti and Clarke, 2008; Hartl et al., 2011; Spiess et al., 2004). Each ring creates a central chamber where substrate polypeptides bind and fold. Unlike simpler archaeal chaperonins, TRiC contains eight different paralogous subunits, named CCT1–CCT8, at fixed positions within each ring (Kalisman et al., 2012; Leitner et al., 2012). All subunits are structural homologs that consist of an ATP-binding equatorial domain and a substrate-binding apical domain linked by an intermediate domain (Bigotti and Clarke, 2008; Spiess et al., 2004) (Figure 1A). Each subunit also contains an apical segment that forms a lid over the cavity. An ATP-driven conformational cycle links TRiC-mediated folding to opening and closure of the lid, encapsulating the substrate in the cavity (Cong et al., 2012; Meyer et al., 2003; Reissmann et al., 2007, 2012).

Understanding how TRiC recognizes its substrates has important implications for human health (Balch et al., 2008). TRiC interacts with approximately 10% of the proteome and is essential for viability (Yam et al., 2008). Mutations in CCT5 and CCT4 are linked to sensory neuropathy (Bouhouche et al., 2006). Cancer-linked proteins p53, von Hippel Lindau tumor suppressor (VHL), and STAT3 are also TRiC substrates (Kasembeli et al., 2014; Trinidad et al., 2013), and mutations in the TRiC-binding sites of VHL lead to misfolding tumorigenesis (Feldman et al., 1999, 2003). TRiC also suppresses aggregation and toxicity of Huntingtin in Huntington's disease (Behrends et al., 2006; Kitamura et al., 2006; Tam et al., 2006, 2009). TRiC is also important for folding viral proteins and required for replication of important human pathogens, including HCV and HIV (Inoue et al., 2011; Zhou et al., 2008). In HIV, TRiC interacts with proteins Gag, Vif, and p6 (Hong et al., 2001; Jäger et al., 2012).

The unique architecture and mechanistic features of TRiC set it apart from other chaperones. The diversification of subunits in TRiC is likely central to understand why many essential proteins, such as actin, Cdc20, and Cdh1, can only be folded with assistance from TRiC (Hartl et al., 2011; Spiess et al., 2004). Despite their extensive conservation in the ATP-binding domains, TRiC subunits have widely divergent functions within the ATP-driven cycle (Reissmann et al., 2012). Additionally, the surface properties of the different subunits result in an asymmetric distribution

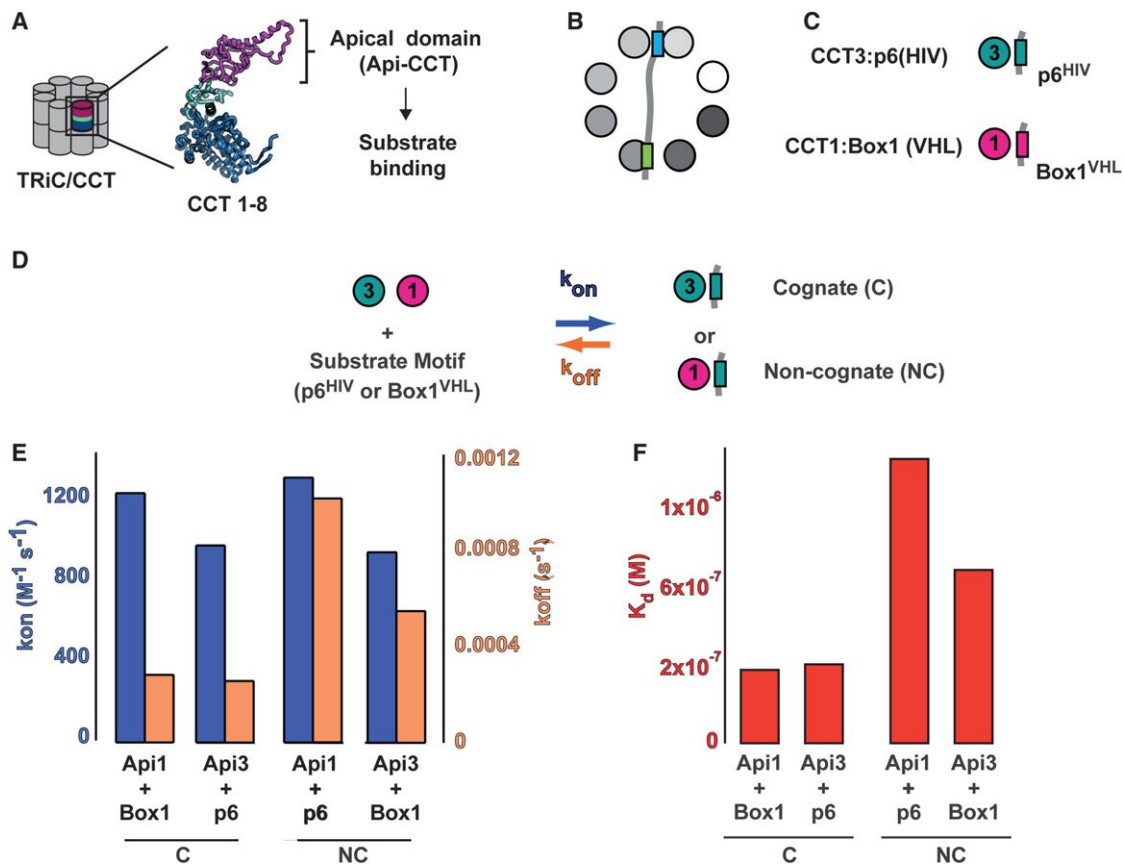


Figure 1. Kinetic Analysis of Substrate Motif Recognition by TRiC Apical Domains

(A) TRiC/CCT subunit domain architecture.

(B) Substrate polypeptides bind multiple TRiC subunits through discrete motifs.

(C) Previously characterized subunit-substrate motif interactions. Top: CCT3 binds retroviral proteins p6 from HIV and p4 from MMPV. Bottom: CCT1 binds Box1, a short linear element from VHL.

(D) Kinetic analysis of cognate (C) and noncognate (NC) interactions between purified apical domains and substrate-derived motifs.

(E) Summary of apparent kinetic parameters for cognate and noncognate interactions, k_{on} , blue bars; k_{off} , orange bars.

(F) Summary of apparent binding constants for cognate and noncognate interactions, K_d , red bars.

of electrostatic charges within the folding chamber (Leitner et al., 2012).

The principles driving TRiC substrate recognition are poorly understood. In vivo, TRiC folds a subset of cellular proteins, suggesting a degree of specificity; however, its substrates are functionally and structurally diverse, indicating the potential to bind a broad array of proteins. The apical domains of each TRiC subunit are thought to recognize different motifs in substrates (Spiess et al., 2004; Spiess et al., 2006) (Figures 1A and 1B). However, to date, no precise structural or sequence rules for TRiC-substrate binding have been identified. We here integrate biophysical and computational structural biology approaches with chemical crosslinking and mass spectrometry (XL-MS) to define the basis of TRiC-substrate recognition. We find that unique subunit-specific patterns of polar and hydrophobic residues underlie the distinct substrate binding properties of each subunit in the complex. The diversification of TRiC subunits thus provides a modular menu of binding specificities that allows for combinatorial recognition of substrate polypeptides. This likely contributes

to TRiC's unique ability to fold structurally diverse and topologically complex substrates. Evolutionary analyses further suggest that diversification of TRiC subunits from its simpler archaeal ancestors enabled the expansion of eukaryotic genomes to acquire proteins with novel folds and functions.

RESULTS

Kinetic Analysis of Substrate Motif Recognition by TRiC Apical Domains

To understand the molecular basis of this recognition specificity, we exploited substrates where the cognate CCT subunit and the relevant substrate motif have been identified (Figure 1C). The 54 amino acid-long HIV protein p6, and the related protein p4 from MMPV, associate directly with subunit CCT3 of TRiC (Hong et al., 2001). A short 6–9 amino acid-long hydrophobic motif in VHL, called Box1, contacts subunit CCT1 (Spiess et al., 2006). Importantly, the isolated recombinant apical domains of each TRiC subunit retain the ability to bind substrates

and substrate-derived motifs with the specificity of the same subunits within the intact complex (Spiess et al., 2006; Tam et al., 2006, 2009).

We used purified HIV-p6 (herein p6) and VHL-Box1 (herein Box1) to examine the association of TRiC apical domains of CCT1 (herein ApiCCT1) and CCT3 (ApiCCT3) (Spiess et al., 2006) with cognate and noncognate substrate-recognition motifs (Figure 1D). A surface plasmon resonance (SPR)-based assay measured association and dissociation kinetics for ApiCCT-substrate pairs (Figures 1E and 1F and Figure S1 available online). Binding kinetics of immobilized VHL-Box1 and HIV-p6 to their cognate and noncognate ApiCCT binding partners were monitored by SPR over a range of concentrations (Figures 1D–1F and S1A–S1D). Apparent association and dissociation rates (Figure 1E) and binding constants (Figure 1F) were calculated from the sensograms (Figure S1). These indicated that the ratio of association over dissociation rates, i.e., the overall affinity, was higher for the cognate ApiCCT-substrate pairs (Figure 1E), consistent with the specificity of these motifs for these subunits within the TRiC complex. The measured on-rates, determined at approximately $10^3 \text{ M}^{-1} \text{ s}^{-1}$, were markedly slower than diffusion-controlled binding (Figure 1E, blue bars) but consistent with the relatively slow substrate-binding kinetics of TRiC (Melki et al., 1997). Cognate interactions exhibited slower dissociation kinetics than noncognate interactions (Figures 1E and S1A–S1D). Both association and dissociation rates contribute to substrate specificity for different subunits. For p6, the difference between cognate and noncognate interaction was largely driven by dissociation rates, whereas for Box1, cognate and noncognate discrimination was a result of differential on- and off-rates (Figure 1E). Of note, even the cognate interactions are relatively weak, with an overall affinity of approximately $0.25\text{--}0.5 \text{ }\mu\text{M}$ (Figure 1F). Accordingly, stable TRiC binding to most substrates will depend on multivalent recognition of several elements in the polypeptide by several subunits in the chaperonin.

NMR Chemical-Shift Mapping of ApiCCT3 Identifies the Substrate-Binding Interface

We focused on the ApiCCT3 and p6 interaction pair to gain a deeper structural understanding of TRiC-substrate recognition. NMR-based chemical shift (CS) mapping was used to identify the substrate-recognition interface in ApiCCT3 (Figures 2 and S2). The ^{15}N - ^1H Heteronuclear Single Quantum Coherence (HSQC) spectrum of ApiCCT3 yielded well-resolved and dispersed spectra, accounting for 142 of 167 peaks, covering 85% of the protein sequence (Figures 2A and S2A and not shown). Standard triple-resonance backbone experiments, guided by specific amino acid labeling to anchor the sequence connectivities allowed us to successfully assign >85% of the peaks in the 2D HSQC spectrum, including all the ApiCCT3 residues perturbed upon substrate addition (Figures S2A–S2D). Titration of increasing amounts of unlabeled p6 into ^{15}N -labeled ApiCCT3 produced concentration-dependent shifts in a specific subset of peaks (Figures 2A and 2B); five peaks were strongly perturbed (>0.2 ppm), and another four peaks were perturbed weakly (>0.1 ppm; Figure 2B). Similar experiments were performed with p6-related protein p4 from M-PMV, which binds CCT3 with lower affinity (Hong et al., 2001; data not shown).

p4 addition affected the same residues in ApiCCT3 as p6 did (data not shown), albeit to a lower extent. In contrast, no perturbations were observed upon addition of Box1 (data not shown).

Given that Y247 in ApiCCT3 (Figures 2A and 2B) was strongly perturbed upon substrate binding, we used ^{19}F -NMR on 3F-tyrosine-labeled ApiCCT3 for an orthogonal assessment of the binding interface (Figures 2C and S2E). 1D ^{19}F -NMR spectra of 3F-tyrosine-labeled ApiCCT3 revealed five discrete peaks, consistent with the five tyrosine residues in ApiCCT3 (Figure 2C). Systematic tyrosine-to-phenylalanine point mutations assigned each peak to unique tyrosine residues (Figure S2E). Upon addition of p6, one of the peaks exhibited a well-defined 0.2 ppm shift. In good agreement with our chemical-shift mapping, the perturbed peak corresponded to the ^{19}F -tyrosine peak of Y247 (Figure 2C).

Structural Model of ApiCCT3 from NMR Backbone Chemical Shifts

Guided by NMR-CS information (Figure 2Di), we used CS-Rosetta and modeling to gain a structural understanding of ApiCCT3 in the substrate-bound conformation (Shen et al., 2009) (see Experimental Procedures and Figures S4D and S4E). The lowest energy models were comparable to the deposited ApiCCT3 structure without substrate (Pappenberger et al., 2002). Of note, our NMR-derived structural model resolved the apical protrusion, not resolved in the ApiCCT3 crystal structure and shown to be intrinsically disordered in a previous NMR study of an archaeal apical domain obtained without substrate (Heller et al., 2004).

Our NMR-derived model provided insight into the conformational dynamics of the apical domain (Figure 2Dii). The regions of higher mobility in the structure included the helical protrusion (Figures 2Dii and 2E, red) and to a lesser extent the flexible loop adjacent to helix 11 (Figures 2Dii and 2E, herein PL for proximal loop). The highly flexible helical protrusion is involved in formation of the closed lid (Heller et al., 2004), but its role in the open chaperonin conformation is not well understood. Interestingly, Y247, whose chemical shift was strongly perturbed by substrate binding, is at the “hinge” between the flexible lid-forming protrusion and the apical domain. Y247 may participate in both substrate recognition and modulating the conformation of the lid protrusion for subsequent release when the closed lid forms.

Mapping the substrate-induced CS-perturbations (Figure 2B) onto the ApiCCT3 structure revealed a continuous and extensive interaction surface spanning three sets of secondary structure elements (Figures 2Diii, 2E, and 2F). The CCT3 substrate interaction interface is primarily defined by a shallow groove formed between helix 11 and the PL and comprises approximately $700 \text{ }\text{\AA}^2$. The core of the substrate-binding site was relatively constrained, consisting of residues on the surface of relatively rigid helix 11 (herein H11) (residues 296–306). The distal portions corresponded to more flexible elements, including the PL adjacent to H11 (residues 223–232) and the hinge connecting to the flexible lid-forming protrusion including Y247. Supporting this analysis, we obtained low-resolution diffracting crystals of the p6-ApiCCT3 complex, which, following model building and refinement, demonstrated an additional density in the same

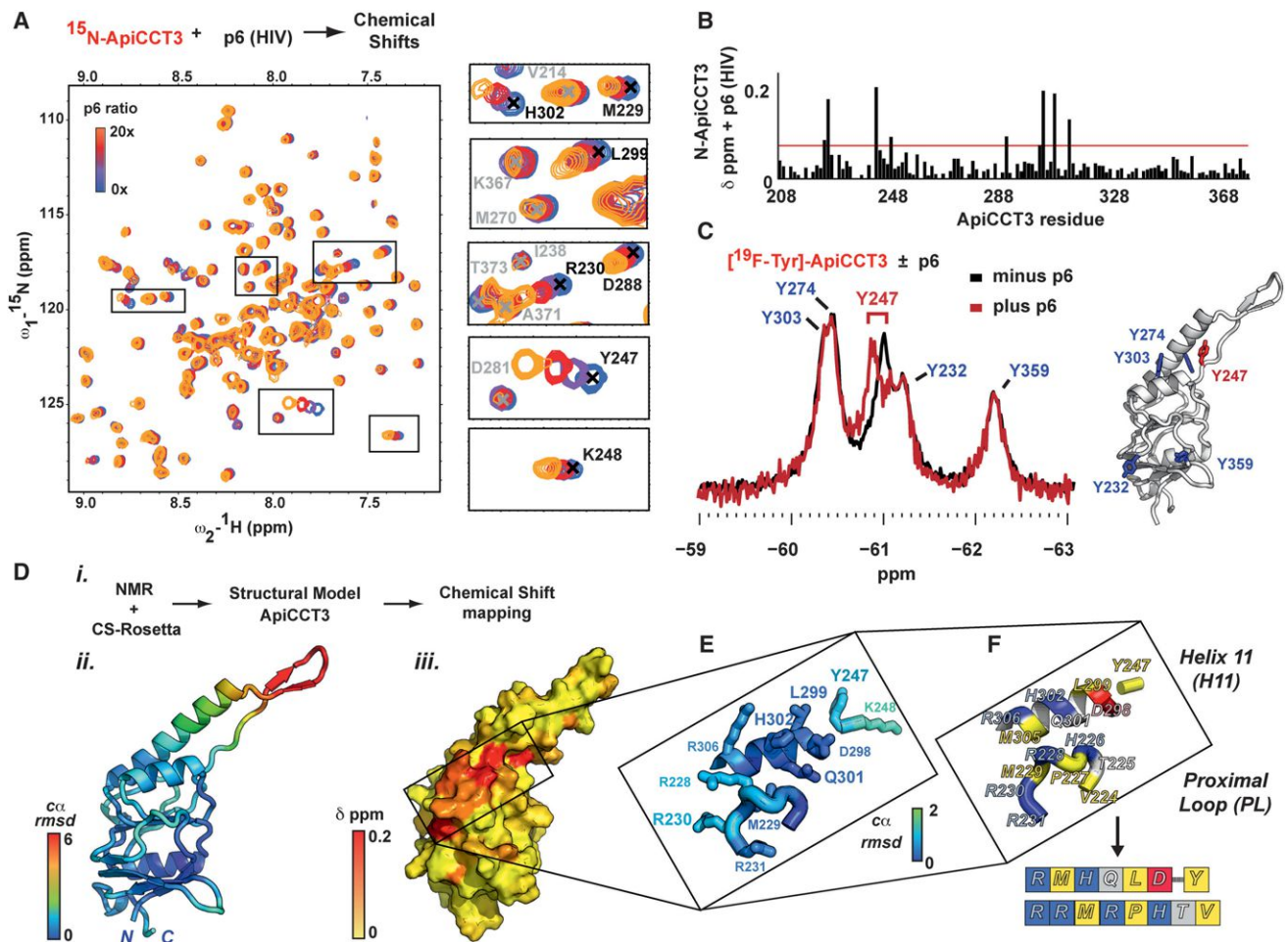


Figure 2. NMR Chemical-Shift Mapping of ApiCCT3 Identifies the Substrate-Binding Interface

(A) 2D NMR spectra of ^1H - ^{15}N ApiCCT3 with increasing amounts of unlabeled p6. Blue, purple, red, and orange, respectively, show perturbations at 0-, 5-, 10-, and 20-fold excess p6 peptide. Insets show perturbed peaks.

(B) Chemical-shift perturbation due to p6 binding mapped to the ApiCCT3 sequence.

(C) ^{19}F tyrosine spectra of ApiCCT3 (black trace). Addition of p6 (red trace) produces a discrete perturbation of Y247. Right panel: ApiCCT3 structure with Y247 in red and the remaining tyrosines in blue.

(D) (i) Integration of NMR chemical-shift parameters with CS-Rosetta. (ii) C- α variability across a subset of low-energy ApiCCT3 models highlights mobile regions. ApiCCT3 shown as cartoon and colored according to C- α rmsd. (iii) p6-induced chemical shifts map to a contiguous ApiCCT3 surface. Surface representation of ApiCCT3 is colored according to the chemical shift from yellow (0) to red (0.2 ppm).

(E and F) Substrate-binding region on ApiCCT3, defined by helix 11 (H11), a loop (PL), and a residue at the hinge of the flexible loop (Y247): (E) Cartoon representation colored according to flexibility (C- α -rmsd as in Dii); (F) colored according to chemical properties of side chains: basic (blue), acidic (red), polar (white/gray), and nonpolar (yellow). Top: cartoon representation of binding site; bottom: simplified box schematic depicting amino acids on the substrate-binding site; upper line: H11 and protrusion hinge residue; lower line: PL sequence.

region of the apical domain identified through NMR (L.A.J., R. McAndrew, J.F., and P. Adams, unpublished data).

Previously characterized chaperone-binding sites, such as those of Hsp70 and the bacterial chaperonin GroEL, rely predominantly on the recognition of hydrophobic determinants (Ashcroft et al., 2002; Chen and Sigler, 1999; Hua et al., 2001; Rüdiger et al., 1997; Swain et al., 2006). In contrast, the substrate-binding site of ApiCCT3 contained a mixture of hydrophobic and polar residues (Figure 2F). In addition to Y247, H11 contributes hydrophobic (L299, M305), polar (Q301), and charged (R306, D298, H302) residues, whereas the PL immediately below

presents a contiguous stretch of basic residues (H226, R228, R230, R231) and a single hydrophobic (M229) residue (Figure 2F, schematically represented by the box diagram in Figure 2F, bottom). Such a shallow, extensive binding surface comprising hydrophobic and polar residues is very different from the mostly hydrophobic substrate-binding sites of Hsp70 and GroEL.

Mutational Analysis Links Chemical Properties of Substrate-Recognition Site to Binding Kinetics

We next designed and purified a large unbiased panel of alanine substitutions in ApiCCT3, comprising 31 surface-exposed

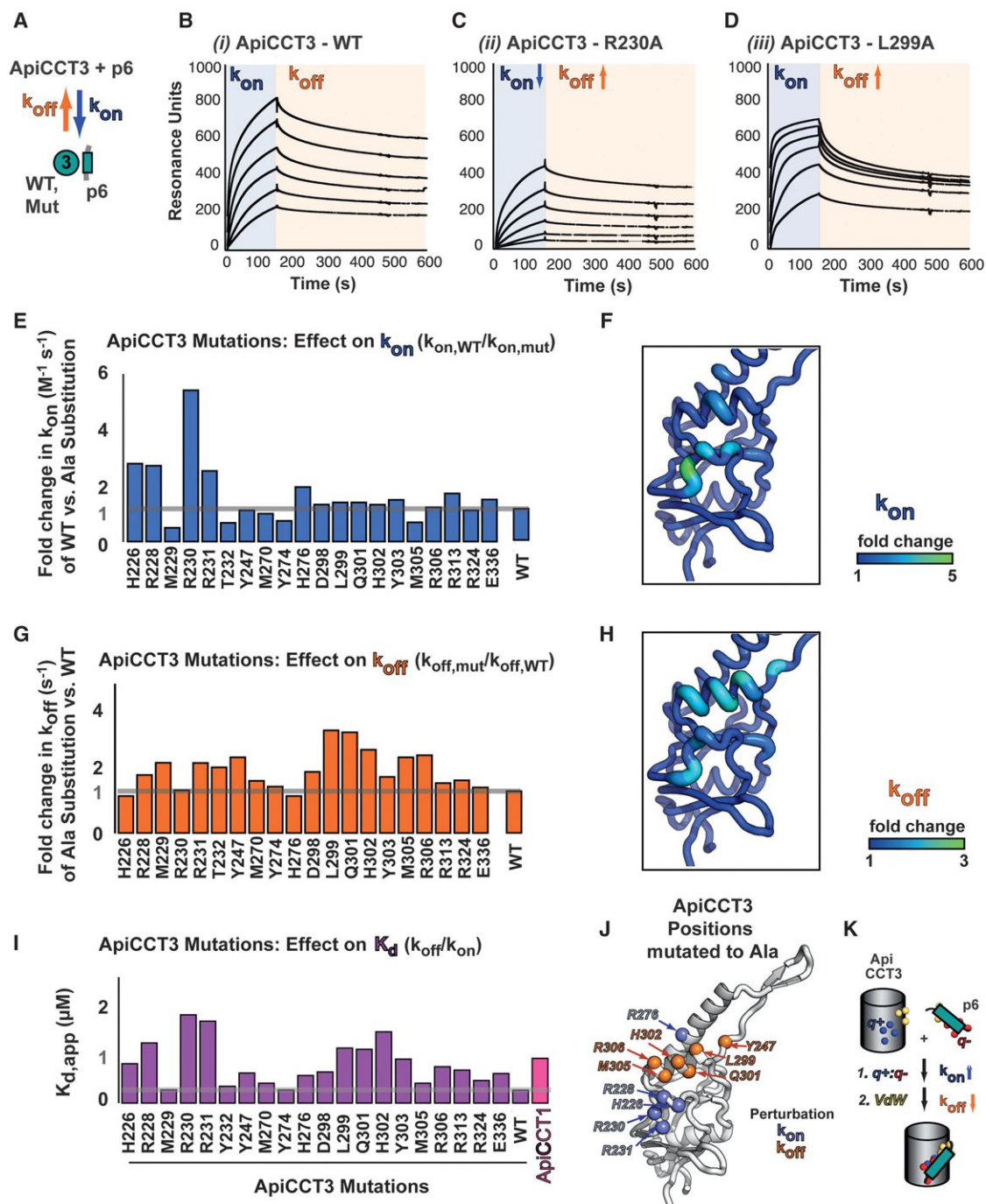


Figure 3. Mutational Analysis of CCT3 Links Chemical Properties of Substrate-Recognition Site to Binding Kinetics

(A) Kinetic scheme of ApiCCT3:p6 interaction (k_{on} in blue, k_{off} in orange).

(B–D) Exemplary SPR titrations of ApiCCT3 variants interacting with an immobilized p6 peptide, performed as in Figure 1. (B) ApiCCT3-WT, (C) ApiCCT3-R230A, and (D) ApiCCT3-L299A. Association phase is in blue, and dissociation phase is in orange. Each titration (black curves) included an analyte dilution series of 50, 25, 12.5, 6.25, and 3.125 μ M apical domain.

(E–H) Summary of fitted kinetic parameters for ApiCCT3 mutants (see also Table S2 for details). (E and G) Bar graphs showing the fold change relative to WT apparent rates for (E) association rate k_{on} (blue, $k_{on,WT}/k_{on,mu}$) and (G) dissociation rate k_{off} (orange, $k_{off,mu}/k_{off,WT}$) from p6 substrate. (F and H) Fold change in rates mapped onto the ApiCCT3 structure highlight the clustering of surfaces based on Ala mutants perturbing binding kinetics through (F) association and (H) dissociation rates. ApiCCT3 structure is colored according to fold change in kinetic contribution; blue is equal to WT, and green is perturbed as indicated.

(I) Calculated apparent K_d for WT ApiCCT3 and each alanine mutant. The apparent K_d of p6 for ApiCCT1, also measured by SPR, is included for comparison.

(legend continued on next page)

residues in the 167 residue apical domain (Figures 3 and S3). The global impact of these mutations on p6 binding was determined with SPR (Figures 3 and S3 and data not shown). Of the initial 31, 10 mutations had no impact on p6 binding and were not examined further, except for a few mutants, such as Y274A, chosen as a control (Figure S3). We determined kinetic binding parameters for 22 alanine mutants, including the one without effect as a control, by carrying out full titration series with SPR, followed by kinetic global fitting of the data (Figures 3B–3D, Table S1, and summarized in Figures 3E, 3G, and S3). This unbiased mutational analysis of ApiCCT3 independently confirmed the NMR-based identification of the substrate-binding site (Figure 2). Thus, mutation of residues perturbed by NMR mapping dramatically affected p6 binding (Figures 3E–3H), whereas mutation at control sites (e.g., Y274 to A274) did not (Figure S3).

Our analysis reveals the kinetic underpinnings of TRiC-substrate recognition (Figures 3F, 3H, and 3I). Substrate association (k_{on}) and dissociation rates (k_{off}) were very differently affected by mutations at discrete positions in the binding site, clustered into two distinct regions (Figures 3F, 3H, and 3I). The association kinetics were predominantly perturbed by mutations of positively charged residues in the flexible PL (Figures 3F and 3I). On the other hand, the contribution to the dissociation rate is distributed across both H11 and the PL. Residues on H11, particularly a mix of nonpolar, polar, and charged side chains, contributed predominantly to the dissociation rate (Figures 3H and 3I). In the PL, the arginine residues allow both charge-charge interactions, likely contributing to the association rates, as well as cation- π and aliphatic chain interactions with nonpolar residues, which likely contribute to the dissociation rates. Thus, the mixed chemical nature of the CCT3-binding site, combining polar and hydrophobic residues, establishes a dual mode of substrate recognition (Figure 3J). The overall contribution of H11 and the PL to the binding constant is distributed across residues L299, Q301, H302 and R228, R230, R231, respectively. These findings resonate with studies of the interfaces between folded proteins, where a core of hydrophobic residues contributes to dissociation rates and polar interactions at the periphery drive association and orientation (Bogan and Thorn, 1998; Clackson et al., 1998).

NMR Identification of the Chaperonin-Binding Determinants in the Substrate

NMR CS-mapping identified next the chaperonin-binding site in the substrate (Figure 4). ^{15}N - ^1H HSQC spectra of ^{15}N -labeled p6 were assigned, and CS-NMR was used to model the peptide structure and identify determinants recognized by ApiCCT3 (Figures 4A and S4A). Adding unlabeled ApiCCT3 to ^{15}N -p6 caused a concentration-dependent chemical shift perturbation in a subset of peaks (Figures 4A and 4B). These peaks mapped to the contiguous and highly conserved S41-N45 element at the p6 C

terminus, consisting of both nonpolar and polar residues (S41-N45; Figure 4C).

To determine the role of the S41-N45 element in chaperonin recognition, we generated and purified the penta-alanine substitution p6_{SLFGN} = > AAAAA (herein p6_{mut}). To determine whether the SLFGN = > AAAAA mutation affected ApiCCT3 binding, we used NMR to examine the perturbation of the ^{15}N -ApiCCT3 spectra upon titration of unlabeled p6_{mut} (Figure 4D). Mutation of the S41-N45 motif largely abrogated the p6-induced chemical shift perturbations in the ApiCCT3 spectrum, indicating that this element mediates chaperonin binding (Figure 4D). We next examined the role of S41-N45 in the interaction of p6 with intact TRiC/CCT (Figure 4E). Purified p6_{WT} or p6_{mut}, labeled with an N-terminal biotin tag, were incubated with mammalian cell extracts, which contain the intact hetero-oligomeric TRiC complex (Figure 4E). Following biotin affinity isolation, the p6 interaction with endogenous TRiC was evaluated by immunoblot analysis (Figure 4E). As expected, p6 bound TRiC, whereas p6_{mut} did not (Figure 4E). These orthogonal approaches support the conclusion that the S41-N45 motif in p6 is indeed the TRiC binding site. To further corroborate this conclusion, we examined the chemical shift perturbations induced upon incubation of ^{15}N p6 with intact purified TRiC (Figure S4D). Indeed the same subset of p6 residues was affected upon addition of intact TRiC, indicating that the ApiCCT3-binding site of p6 mediates its interaction with TRiC.

CS-Rosetta was next used to derive the solution structure of p6 (Figures 4Fi, S4B, and S4C). Analysis of the top-scoring models showed that p6 in aqueous solution contains a structured helical element at the C terminus and a flexible polar region (Figure 4G). The C terminus of p6, containing S41-N45, adopts a helical conformation (Figure 4Fi). The ApiCCT3-interacting residues L42, F43, and N45 map to one face of the helix (Figure 4Fii). In contrast, the N terminus is highly dynamic as highlighted by the $\text{C}\alpha$ -rmsd map (Figures 4Fi, S4B, and S4C). Circular dichroism (CD) measurements revealed only a very subtle decrease in helicity when p6 binds the chaperonin, consistent with the weak helicity observed in the NMR-guided structural model of p6 (Figures S4B and S4E). Interestingly, whereas the helical chaperonin-binding determinant has significant hydrophobic character, the flexible N terminus contains a series of acidic residues (Figure 4G). The structural model of p6 implies that these acidic residues may interact with the basic residues in the ApiCCT3 loop, providing a molecular rationale for the electrostatic-driven association kinetics observed by SPR.

Structural Model of the Chaperonin-Substrate Interface

RosettaDock with CS-derived site constraints was employed to obtain a structural model of the ApiCCT3-p6 complex (Figures 5 and S5). The lowest energy model was fully consistent with the NMR data of both p6 and ApiCCT3 (Figures 5A and 5B), and

(J) Residues that contribute to association and dissociation rates are shown as spheres on a cartoon model of ApiCCT3; perturbations affecting association rates (blue) cluster in the PL; those affecting dissociation rates (orange) cluster on H11; and the hinge is at the apical protrusion. Inset, surface representation of ApiCCT3-p6-binding site colored according to amino acid chemical property: basic in blue, acidic in red, polar in white, and nonpolar in yellow.

(K) Differential kinetic contribution of two regions in ApiCCT3 to substrate binding: charge-charge interactions between basic and acidic residues in blue and red, respectively, control association rates (R228, R230, R231 also contribute to off-rates), whereas a mix of nonpolar, polar, and Van der Waals interactions, shown in yellow, control dissociation rates.

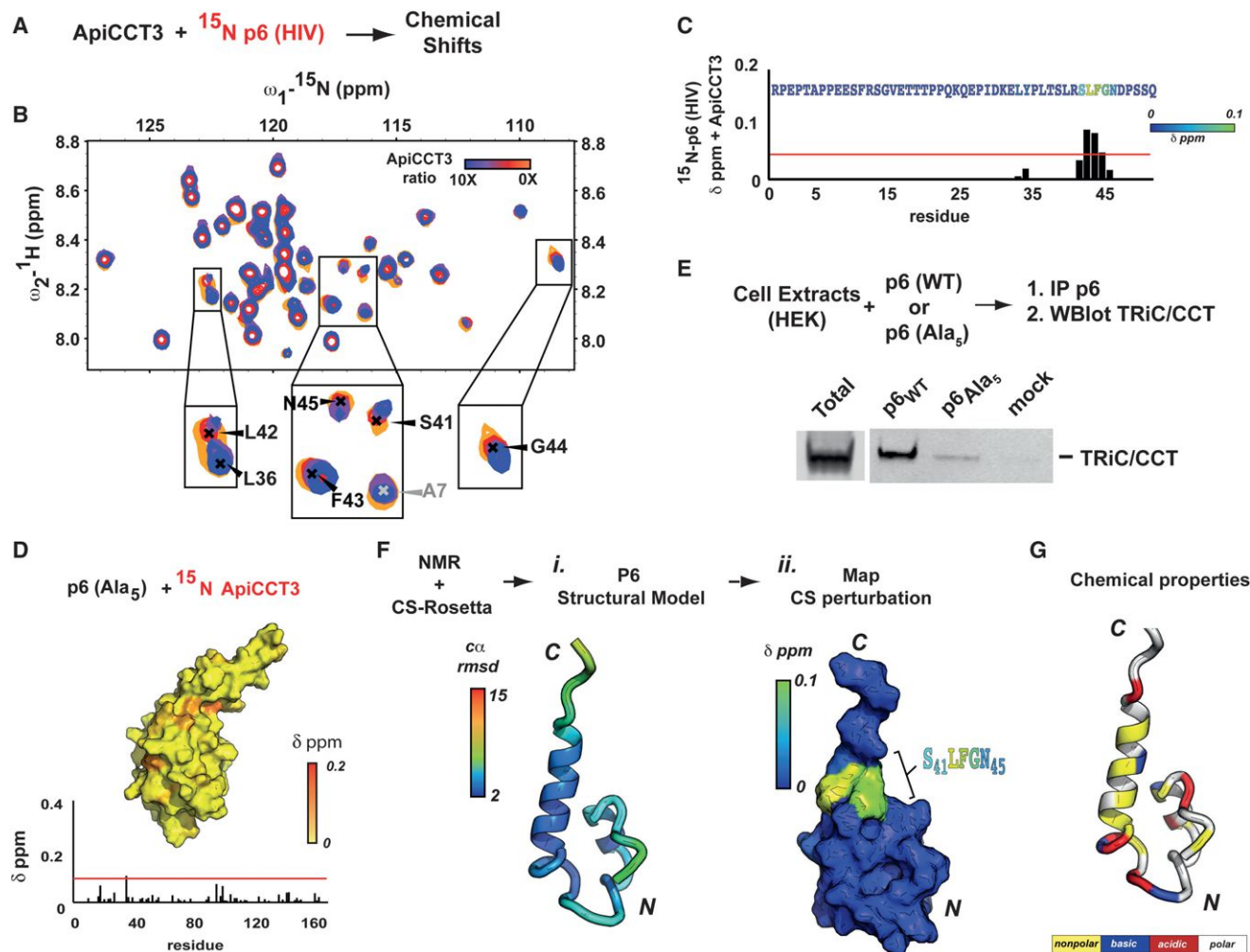


Figure 4. NMR Identification of the Chaperonin-Binding Determinants in the Substrate

(A) NMR-based chemical-shift mapping strategy identifies the p6-interacting surface.

(B) 2D NMR spectra of ^1H - ^{15}N p6 with increasing amounts of unlabeled ApiCCT3. Orange, red, purple, and blue, respectively, show perturbations at 0-, 2.5-, 5-, and 10-fold excess ApiCCT3.

(C) Chemical-shift perturbation mapped onto the p6 sequence colored by ApiCCT3 perturbation.

(D) Mutant p6_{mut} (p6_{SLFGN} > AAAAA) does not induce chemical-shift perturbations in the ^{15}N ApiCCT3 spectrum observed with p6_{WT}. Perturbations mapped on the ApiCCT3 sequence and on the ApiCCT3 surface, exactly as in Figure 2F.

(E) p6_{mut} mutation abrogates the binding to endogenous TRiC observed for p6_{WT}. Biotinylated p6_{WT} or p6_{mut} was incubated with extracts from human HEK293 cells, affinity isolated via the biotin tag (IP). TRiC in the IP is visualized by western blot. Total: 1% input to the IP.

(F) Structural model of p6 from CS-NMR. (i) C- α variability of p6: cartoon representation according to variability across a subset of low-energy models, as indicated. (ii) Surface representation of p6 colored according to chemical-shift perturbation by ApiCCT3 addition, as indicated.

(G) Chemical properties of p6, colored according to amino acid properties: blue is basic, red is acidic, white is polar, and yellow is nonpolar.

the interface agreed with all our experimental data, including CS perturbations (Figures 5C, 1, and 3) and mutagenesis (Figures 2, 3, and 5D).

The ApiCCT3-p6 structure provides unprecedented detail on chaperonin-substrate binding (Figures 5A and 5B). p6 makes tight packing interactions with unique features in the ApiCCT3 H11 and PL region through the specific presentation of side chains (Figures 5A and 5B). The interface consists of two distinct regions, highlighting the dual nonpolar and polar nature of binding and providing a rationale for the bipartite substrate-binding mode observed in the kinetic analyses. The interaction core is

established by a mix of nonpolar and polar interactions, centered on H11 (Figures 5B–5D; namely L299, H302, M305, Q301, and Y247 in ApiCCT3 and L42, F43, and N45 in p6). A region of mostly electrostatic interactions is centered primarily on the PL, between positively charged residues in ApiCCT3 and acidic residues in p6 (Figure 5B). These charge-charge interactions both confer specificity for discrete elements in the substrate and serve to orient the substrate upon binding the apical domain (Figure 5B). Additional nonpolar contacts make close packing interactions with the aliphatic chains of lysine and arginine in the apical domain. As a result, of the 1073 Å² of buried surface

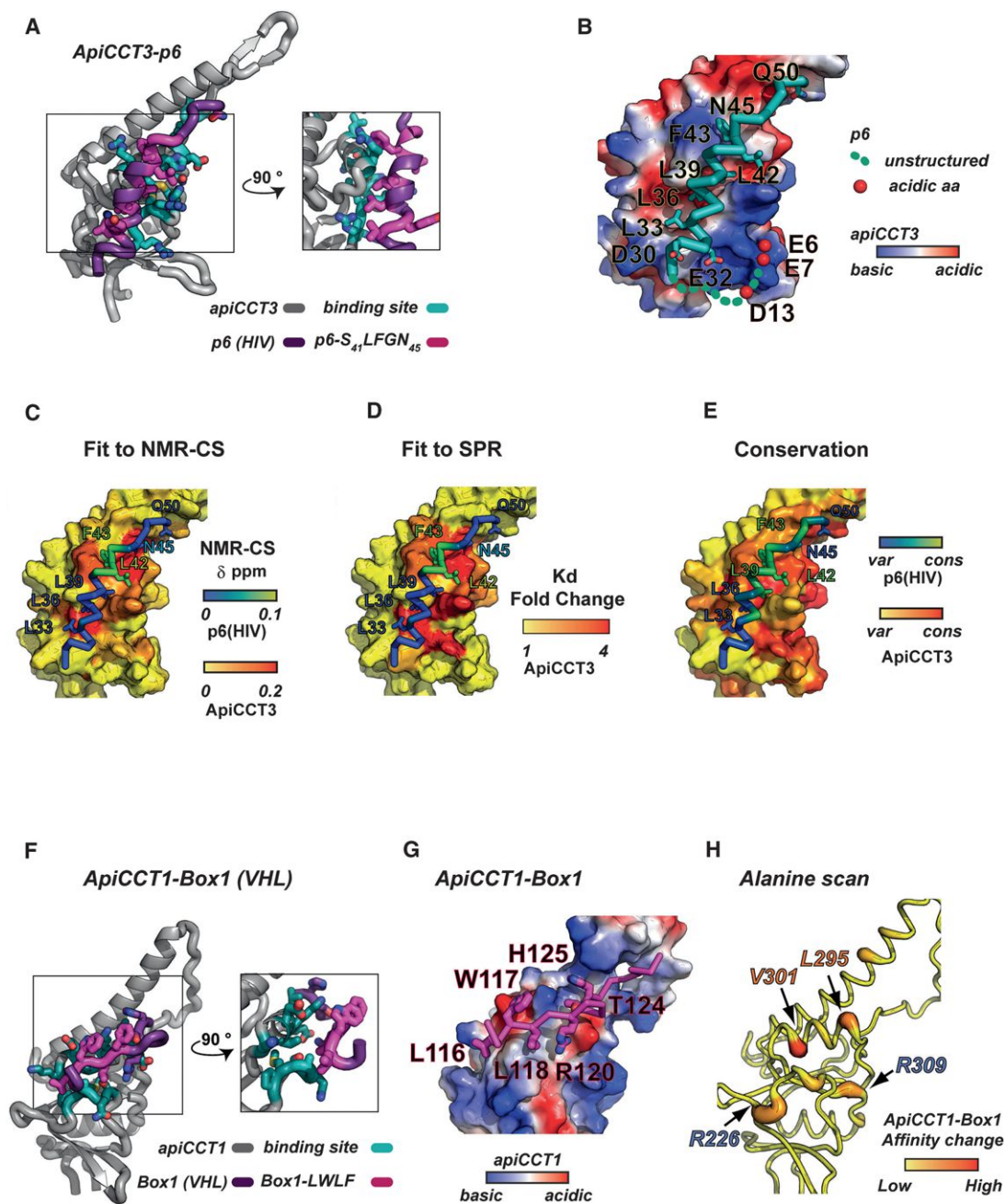


Figure 5. Structural Model of the Chaperonin-Substrate Interface

(A and B) Structural model of *ApiCCT3-p6*. (A) *ApiCCT3* and *p6* colored in gray and purple, respectively. Interfacial residues in stick representation are colored in teal and magenta. Inset, zoom-in of the interface at 90° rotation. (B) *ApiCCT3* substrate-binding interface in surface representation colored according to electrostatic potential; blue, red, and white are positive, negative, and neutral residues, respectively. Bound *p6* is shown in teal with interfacial residues in stick representation.

(C–E) Agreement between experimental data and structural model of *ApiCCT3* (surface) and *p6* (ribbon): (C) Reciprocal chemical-shift perturbations in *p6* (blue-green) and *ApiCCT3* (yellow-red) upon complex formation. (D) Mutagenesis perturbation of binding measured by apparent K_d . (E) Evolutionary conservation among orthologs of *CCT3* (yellow-red) and *p6* (blue-green).

(F–H) Structural model of *ApiCCT1-Box1*. (F) *ApiCCT1-Box1* complex in cartoon representation colored in gray and purple, respectively. Interfacial residues shown as sticks and colored in teal and magenta. Inset, zoom in of the interface at 90° rotation. (G) Electrostatic charge distribution of *ApiCCT1* substrate-binding site bound to the *Box* peptide. Surface representation of the *ApiCCT1* substrate-binding site colored according to electrostatic charge potential; blue, red, and white are positively, negatively, and neutrally charged residues, respectively. Bound *Box1* in magenta with interfacial residues shown in stick representation. (H) Mutagenesis analysis of *CCT1* residues required for *Box1* binding. Affinity-ranking SPR experiments assessed binding of alanine *ApiCCT1* mutants for *Box1*. *ApiCCT1* residues that perturb binding are highlighted in red/orange on the cartoon putty representation.

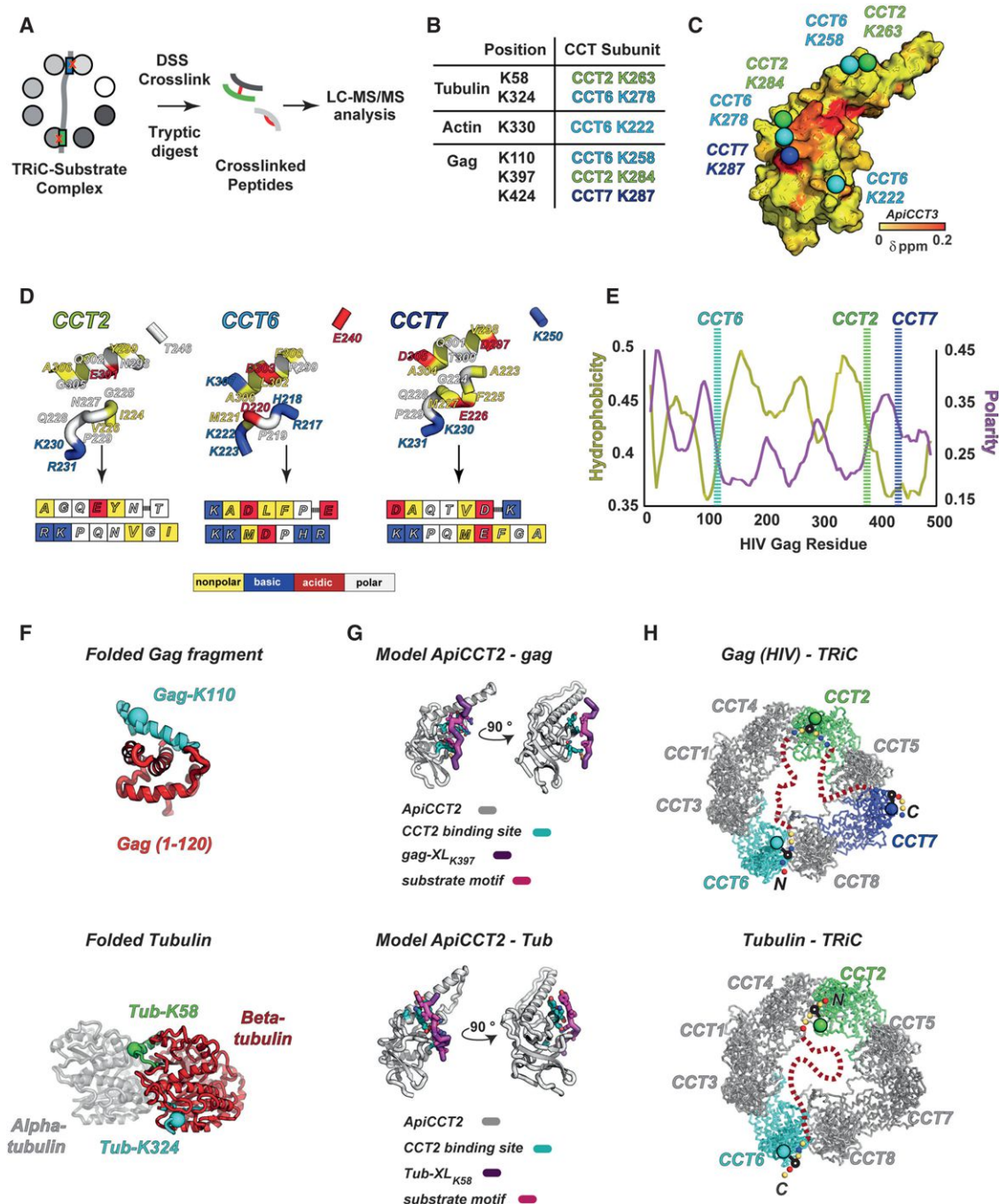


Figure 6. Mapping the Contacts between the TRiC Hetero-oligomer and Full-Length Substrates

(A) XL-MS approach to map TRiC-substrate contacts.

(B) XL-MS-derived contact points between tubulin, actin, and Gag and specific sites on TRiC.

(C) Substrate crosslink sites mapped onto the surface representation of ApiCCT3. The substrate-binding interface in ApiCCT3 is highlighted as in Figure 2F. The sites of substrate crosslinks shown as spheres are colored green, cyan, and blue for CCT2, CCT6, and CCT7, respectively.

(D) Comparison of H11/PL substrate-binding region of subunits CCT2, CCT6, and CCT7 colored according to amino acid properties; blue, red, and white correspond to basic, acidic, and neutral amino acids. Bottom: schematic representation of binding region in box format as in Figure 2H, colored as indicated.

(E) TRiC crosslink sites mapped onto the Gag protein sequence. Vertical lines show sites where Gag crosslinks to the apical domains of CCT6 (cyan), CCT2 (green), and CCT7 (blue). The plot also maps the hydrophobicity (yellow) and polarity (purple) of Gag.

(F) Gag and tubulin crosslinks to TRiC mapped onto the folded substrate structures. (Top) Cartoon representation of N-terminal domain of Gag in red with flexible region containing the site of CCT6 crosslink (K110, as spacefill) in cyan. (Bottom) Cartoon representation of tubulin heterodimer: β-tubulin in red and α-tubulin in gray. Loops containing the two β-tubulin-TRiC crosslink sites (in spacefill: K58 in green to CCT2; K324 in cyan to CCT6) are shown in green and cyan.

(legend continued on next page)

area, 762 Å² correspond to nonpolar contacts and 311 Å² to polar contacts.

Integrating conservation across orthologs for ApiCCT3 and p6 sheds light on the potential coevolution of surfaces employed in chaperone-substrate interaction (Figure 5E). The core interface residues are conserved in both ApiCCT3 as well as p6 variants across HIV clades. Conservation in this p6 region could respond to the requirement for TRiC interaction and/or interaction with orthogonal binding partners such as VPR and ALIX, which also bind in this p6 region (Salgado et al., 2009).

We next employed a similar RosettaDock-based analysis to obtain a structural model for VHL-Box1 in a complex with the apical domain of CCT1 (Figures 5F, S5A, and S5B). The ApiCCT1-Box1 structural model placed Box1 at the same H11/PL region of ApiCCT1 where p6 binds ApiCCT3 (Figure 5F). Box1 adopts an extended conformation upon binding (Figure 5G). The side chains of L116 and W117 in Box1, known to be critical for VHL binding to TRiC *in vivo* and *in vitro* (Feldman et al., 2003), pack between H11 and the PL in ApiCCT1 (Figure 5F, inset and Figure 5G). Interestingly, comparison of the substrate complexes of CCT1 and CCT3 (Figure 5B versus Figure 5G) shows that the substrate-binding surface of CCT1 is more hydrophobic than that of CCT3, consistent with the higher hydrophobicity of Box1 over p6. We validated the ApiCCT1-Box1 structural model using ApiCCT1 alanine-substitution mutagenesis followed by affinity measurements (Figures 5H and S5C). A set of 25 mutants in ApiCCT1 were purified, and their interaction with Box1 analyzed by an affinity-ranking SPR approach (Figures 5H and S5C). Strikingly, the five ApiCCT1 alanine mutants that most significantly perturbed VHL-Box1 binding mapped to the interface predicted by the structure (Figure S5C). These data, together with previous analysis of Box1 residues required for TRiC binding (Feldman et al., 2003), validate the structural model for ApiCCT1-Box1. We conclude that the groove formed between H11 and the flexible PL is the general substrate-recognition site of TRiC/CCT subunits.

Mapping the Contacts between the TRiC Hetero-oligomer and Full-Length Substrates

To extend our understanding of TRiC substrate recognition to full-length substrates, we used chemical crosslinking-mass spectrometry (XL-MS) to identify contact points between intact TRiC and three full-length physiological substrates: actin, tubulin, and HIV Gag (Figures 6 and S6). For actin and tubulin, the heterotypic TRiC crosslinks localized to substrate determinants previously implicated in TRiC binding by peptide arrays and mutagenesis (Table S2 and Figure 6B) (Hynes and Willison, 2000; Ritco-Vonsovici and Willison, 2000; Rommelaere et al., 1999). Both tubulin and Gag crosslinked to multiple TRiC subunits through specific regions in the polypeptide (Table S2 and

Figure 6B), consistent with a multivalent contact between TRiC and its substrates.

Exploiting the conservation of general architecture of TRiC subunits, we mapped the location of substrate crosslinks to CCT2, CCT6, and CCT7 (Figure 6C, green, blue, and cyan spheres, respectively) onto the ApiCCT3 structure, highlighting its substrate-binding surface (Figure 6C, red surface). Strikingly, the substrate crosslinks are proximal to the apical domain substrate-binding interface between CCT3-p6 and CCT1-Box1, validating this region as the general location of the substrate-binding site in all TRiC subunits.

Comparing the chemical properties of the H11/PL region in subunits CCT2, CCT6, and CCT7 (Figure 6D) shows that the substrate-binding site of each subunit has a distinct pattern of hydrophobic and polar residues. Thus, the dual-recognition mode observed for CCT3-p6 is a general feature of TRiC-substrate recognition. Interestingly, analysis of the location of crosslink sites in the substrate primary sequence (Figure 6E for Gag; see also below; Figure S6C) indicated that chaperonin contact points within the polypeptide are close to the boundary between a nonpolar and polar region (Figure 6E, yellow trace: hydrophobicity; purple trace: polarity). The distinctive combination of polar and hydrophobic elements in both the substrate and each chaperonin subunit H11/PL region may underlie subunit-specific interactions (Figures 6 and S6C).

We next mapped the TRiC-crosslink sites onto the folded structures of actin, tubulin, and Gag (Figure 6F). The chaperonin-contact sites are proximal to both a structured hydrophobic region, either helix or strand, and a more unstructured polar loop (Figure 6F, top panel, N terminus of Gag; bottom panel, tubulin, not shown for actin). The tubulin crosslinks map to two surface loops at the tips of the N-terminal and C-terminal lobes of the protein (Figure 6F, bottom panel). The TRiC-binding sites overlap with the interface of the tubulin heterodimer, indicating that folded and assembled β -tubulin cannot bind to TRiC (Figure 6F, bottom). The binding site for tubulin assembly factor Rbl2/Cofactor A (CoA), which acts directly downstream of TRiC in tubulin assembly (Tian and Cowan, 2013), also overlaps with the Tub2 site of crosslink to CCT6 (You et al., 2004). The overlapping tubulin-binding sites for TRiC and CoA suggest possible mechanisms for Tub2 release from TRiC and indicate that the chaperonin protects this oligomerization surface from inappropriate intra- and intermolecular interactions.

To understand how the same subunit can bind distinct substrate motifs, we used the crosslinking information as a physical constraint to generate models of the CCT2-substrate interaction with Gag and tubulin (Figures 6G and S6A–SC). The lowest energy models placed both substrate-derived peptides in the CCT2-binding site formed by H11 and the PL, even though the starting distance constraint, i.e., the site of crosslink, was

(G) XL-MS-derived structural models of CCT2 apical domain-substrate interaction with substrate elements from crosslink sites of Gag (top) and tubulin (bottom). Apical domain-peptide complexes are colored in gray and purple, respectively. Interfacial residues in stick representation are colored teal for the apical protein and magenta for the peptide.

(H) Putative topological description of the TRiC-bound substrate (dashed line) for Gag (top) and tubulin (bottom). The open state of TRiC is shown in gray, with subunits CCT2, CCT6, and CCT7 colored green, cyan, and blue, respectively. The C α lysines involved in the crosslink are shown as spheres. The immediate proximity of the crosslink site on the peptide sequence is colored according to amino acid properties; yellow, white, blue, and red are nonpolar, polar, basic, and red, respectively.

distal from this site. Gag and tubulin bind in different configurations to the same apical domain of CCT2. Comparing all the data and structural models obtained here for different apical domain-substrate complexes reveals common rules for TRiC recognition and specificity. The shallow groove created by H11 and the PL allows flexibility in binding, allowing the same apical domains to bind different substrates with no sequence similarity (Yam et al., 2008). H11 and the PL provide the apical domain surface for substrate recognition through a combination of polar and hydrophobic interactions. The specific polar-hydrophobic pattern of both apical domain and substrate serve to provide specificity and orient the substrate to dictate the binding topology.

Most excitingly, the XL-MS analysis provided a topological description of the substrate when bound to TRiC (Figure 6H). Subunit-specific contacts provide anchors that determine a global configuration of substrate polypeptides bound to TRiC; for Gag, the polypeptide is stretched across the open complex, whereas for tubulin, the two contact points at the tips of its two lobes are at antipodal positions of the ring. Taken together, these data indicate that binding to TRiC orients and restricts the global topology of the bound substrate, perhaps allowing the domains to start folding while associated with the chaperonin.

DISCUSSION

How TRiC/CCT discriminates between non-native substrates and their folded counterparts is intriguing in view of its obligate requirement for folding a subset of cellular proteins that share no sequence or structural similarities. By defining the structural basis of substrate recognition, we begin to understand how subunit diversification enabled TRiC to balance the plasticity required to recognize a broad array of substrates with the specificity required to assist their folding.

Implications for Substrate Selection and Folding in the Hetero-oligomeric Chaperonins

Mapping the substrate binding in the apical domains of TRiC to a groove between H11 and the PL has fundamental implications for the mechanics of substrate folding (Figure 7A). Within the intact complex, the interaction with each subunit relies on a recognition code integrating polar and hydrophobic contributions, which enables combinatorial substrate recognition (Figure 7Ai). Polar contacts orient the substrate locally upon binding, and the distribution of subunit-specific substrate interactions stipulates the global topology of the TRiC-bound polypeptide. This may direct folding of TRiC-bound substrates along a preferred pathway (Figure 7Aii). Our mapping of the substrate-binding site provides a compelling mechanism of substrate release upon ATP-dependent closure (Douglas et al., 2011) (Figure 7Aiv). ATP-induced closure of the lid brings the PL region in one apical domain into direct contact with a loop in the neighboring subunit, termed RLS (release loop of substrate) (Figure 7Aiii), which mediates substrate release locally, through ATP-induced contacts between adjacent apical domains (Douglas et al., 2011) (Figure 7Aiv). This mechanism of release is well suited to the low affinity of each individual apical domain-substrate interaction, as it permits the local displacement of the

substrate from the apical domain by the ATP-induced proximity of the RLS. Because ATP binding and hydrolysis function within the ring is asymmetric (Reissmann et al., 2012) (Figure 7A, dark gray: high ATP affinity; light gray: low ATP affinity), it is possible that the substrate is released sequentially during the conformational cycle (Figure 7A, brackets). The particular dissociation rates for a given subunit will determine when specific regions of the polypeptide are released from their binding sites into the folding chamber. By allowing certain subdomains to fold first, TRiC may promote productive folding trajectories.

Principles Driving the Diversification of Recognition and Specificity in TRiC Subunits

NMR, mutagenesis, modeling, and XL-MS indicate that all TRiC subunits contact substrates through the same region in their apical domains. This region is evolutionarily conserved across orthologs (i.e., across all CCT5 from eukaryotes, Figures 7C, S7A, and S7B) but diverges across TRiC/CCT paralogs (Figure 7B), suggesting a distinct and important function in each subunit.

Chaperonin-binding sites balance plasticity and specificity in substrate recognition through two modular elements that discretely control binding and specificity: charged and polar residues contribute to enhancing on-rates and hydrophobic residues contributing to decreased off-rates. The polypeptide-binding platform in the H11/PL region combines a rigid helical element and a flexible loop. Substrate interaction involves burial of nonpolar residues in the groove formed by the H11 helix/PL loop region. The PL loop is highly variable among subunits, in terms of both chemical properties and length (Figures S7B and 7C). Providing conformational flexibility to PL in the polypeptide-binding groove may enable recognition of a larger set of substrates. Among all subunits, CCT2 presents the most nonpolar, classical “chaperone-like,” binding surface (Figure 7C); interestingly this subunit is crosslinked to all full-length substrates examined.

The shallow nature of the chaperonin-binding groove allows a subunit to recognize motifs with different features, providing plasticity in binding, as shown for CCT2 (Figure 6). The low affinity for a single substrate-apical domain interaction is consistent with an avidity-driven interaction, whereby multiple discrete low-affinity contacts to different subunits mediate stable binding to the complex (Figure 7A). This combinatorial recognition suggests a simple model for the discrimination of folded from nonfolded proteins. TRiC will recognize those conformations interacting with more than one subunit but will not interact with proteins where most binding motifs are no longer available.

The Evolution of Hetero-oligomeric Chaperonins

It is intriguing to consider what drove the evolution of such a complex hetero-oligomeric folding machine. TRiC substrates tend to encode complex topologies, and many coevolved with TRiC to the point of being unable to fold in its absence, (e.g., actin). Archaea have simpler chaperonins, ranging from one to five subunits depending on the organism (Bigotti and Clarke, 2008). We considered whether changes in the proteome are linked to subunit diversification. Strikingly, comparing all organisms containing TRiC-like chaperonins, we find a positive

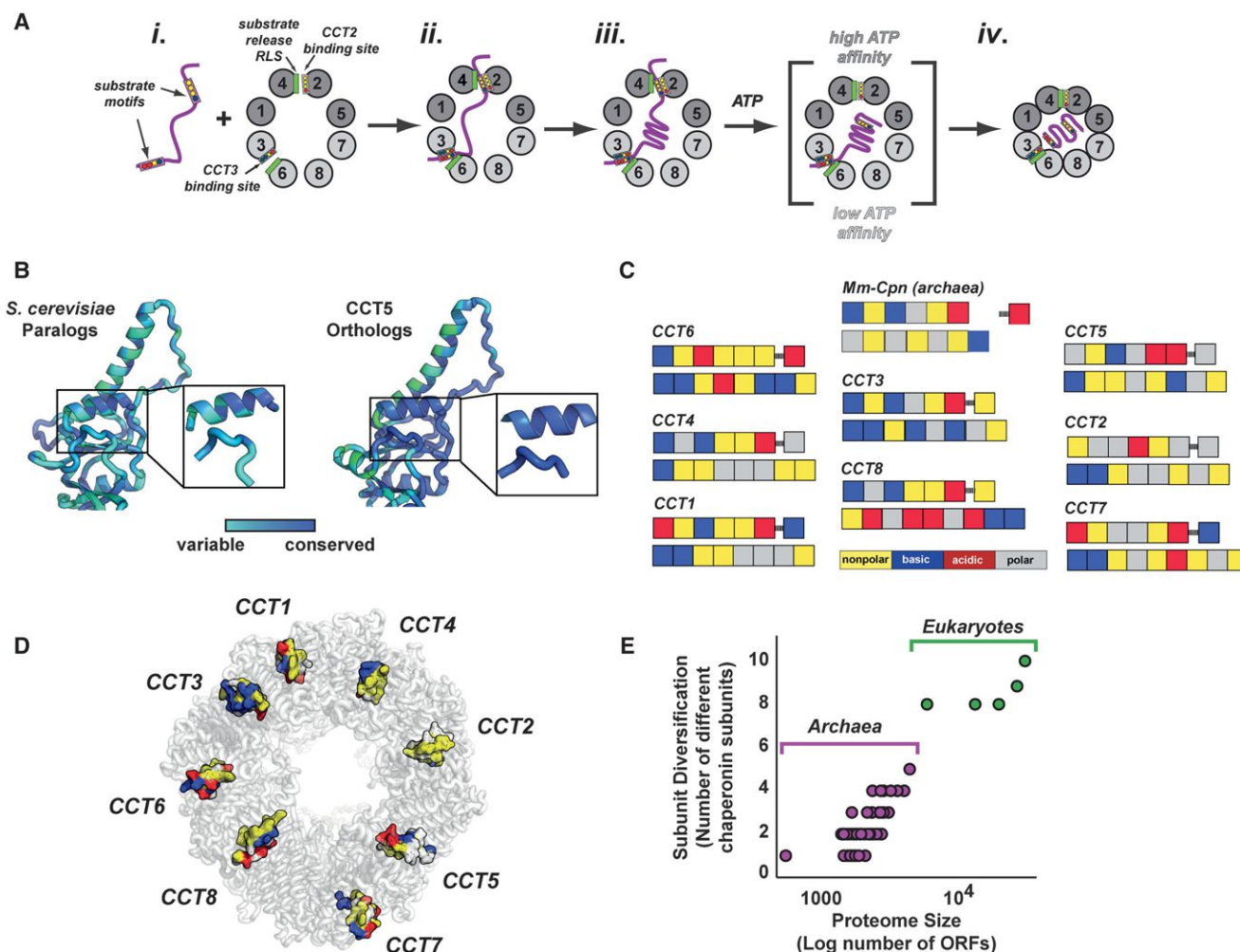


Figure 7. The Polypeptide-Binding Site of TRiC/CCT: Functional and Evolutionary Implications

(A) Role of subunit-specific substrate recognition in the context of the TRiC folding cycle. See Discussion for description. Subunits with high ATP affinity are in dark gray, with low ATP affinity in light gray.

(B) Evolutionary divergence across H11/PL substrate-binding sites across *S. cerevisiae* paralogs (left) and conservation of binding site across orthologs (CCT5, right panel).

(C) Substrate-binding-site properties of the different yeast TRiC subunits schematically shown in box format as in Figure 2H. Upper line corresponds to H11/apical hinge residues, and lower line corresponds to PL. The substrate-binding site of archaeal chaperonin from *M. maripaludis* is included for comparison.

(D) The substrate-binding-site interface of the open TRiC conformation. TRiC is shown in semi-transparent gray cartoon, and substrate-binding sites in surface representation colored according to chemical properties, as indicated.

(E) Group II chaperonin subunit complexity increases with proteome size. Number of subunits in archaeal (purple) and eukaryotic (green) organisms graphed versus proteome size, as number of open reading frames (ORFs) in that organism.

correlation between subunit diversity and the size of its proteome (Figure 7E). The possible link between subunit diversification of TRiC and expansion of the proteome in eukaryotes raises questions on the mechanisms linking protein evolution to changes in chaperone structure and composition. Subunit diversification may increase the probability that a given protein will present two or more binding sites that can combinatorially bind to the chaperonin and benefit from the mechanistic advantages of folding in its chamber. Thus, the complexity of the chaperonin appears functionally optimized for the complexity of the genome, suggesting that the folding machinery contributes to dictate pro-

teome size. A better understanding of the substrates and recognition motifs for the different TRiC and archaeal subunits may provide exciting insights into protein evolution.

EXPERIMENTAL PROCEDURES

Apical domains of TRiC were expressed and purified as described previously (Spiess et al., 2006). A set of alanine mutants were cloned and purified based on the chemical-shift perturbation experiments, and binding kinetics were analyzed by SPR. All apical domains were soluble and folded, as assessed by CD. For the SPR experiments, peptides were immobilized using maleimide chemistry on a PEG-derivatized surface, and a dilution series of apical protein

was flowed over as analyte. For NMR, $^{15}\text{N}/^{13}\text{C}/^2\text{D}$ ApiCCT3 and $^{15}\text{N}/^{13}\text{C}$ p6 samples were expressed and purified using standard isotope-labeling procedures (see Extended Experimental Procedures for details). NMR chemical-shift mapping and backbone assignment experiments for ApiCCT3 and p6 were carried out using 300 μM protein sample on a 800 MHz Inova Varian spectrometer outfitted with a cryogenic probe. Biotinylated p6_{WT} or p6_{mut} peptides were incubated with extracts from human HEK293 cells and affinity isolated via the biotin tag (IP). The presence of TrIC in the IP is visualized by western blot analysis against TrIC antibodies raised against CCT5 and CCT2. DSS-crosslinked TrIC-substrate samples were treated with trypsin, enriched for crosslinked peptides by size-exclusion chromatography and analyzed by tandem mass spectrometry. Crosslinked peptides were identified by xQuest (Rinner et al., 2008). Conservation scores were calculated using Rate4site (Pupko et al., 2002) and mapped onto the models using ConSurf (Ashkenazy et al., 2010). Generation of structural models was guided by experimentally determined backbone chemical-shift parameters in CS-Rosetta (Lange et al., 2012; Mao et al., 2014; Shen et al., 2009; van der Schot et al., 2013). RosettaDock was used to model the ApiCCT-substrate complexes with CS-NMR-based site constraints or with XL-MS-based atom pair constraints. Homology model building for the remaining ApiCCT domains was carried out using the Rosetta software package (Chivian and Baker, 2006; Davis and Baker, 2009).

SUPPLEMENTAL INFORMATION

Supplemental Information includes Extended Experimental Procedures, seven figures, and four tables and can be found with this article online at <http://dx.doi.org/10.1016/j.cell.2014.10.042>.

AUTHOR CONTRIBUTIONS

L.A.J. and J.F. conceived the project and designed and interpreted experiments. L.A.J. cloned, expressed, and purified all apical domain proteins, performed SPR experiments and modeling calculations, and prepared the TrIC-substrate complexes for XL-MS. L.A.J. and C.W.L. performed all NMR experiments. The 800 MHz Agilent VNMRs console was purchased with funds from NIH Shared Instrumentation Grant 1 S10 RR025612-01A1. T.W. performed XL-MS experiments; T.W. and R.A. carried out the XL-MS analysis and interpreted the data. L.A.J. and J.F. wrote the manuscript. All authors contributed to the preparation of the manuscript.

ACKNOWLEDGMENTS

We thank Drs. R. Andino, A. Joachimiak, D. Gestaut, and K. Dalton for comments on the manuscript and Drs. L. Gierasch and D. Agard for advice. We thank Dr. M. Eckart from the PAN facility for help with the SPR experiments. This work was supported by an NIH fellowship to L.A.J., NIH grants to J.F., ERC grant to R.A. (Proteomics v3.0; AdG-233226 to R.A.), and funds from ETH Zurich to R.A. and T.W.

Received: June 22, 2014

Revised: August 17, 2014

Accepted: October 20, 2014

Published: November 20, 2014

REFERENCES

- Ashcroft, A.E., Brinker, A., Coyle, J.E., Weber, F., Kaiser, M., Moroder, L., Parsons, M.R., Jager, J., Hartl, U.F., Hayer-Hartl, M., and Radford, S.E. (2002). Structural plasticity and noncovalent substrate binding in the GroEL apical domain. A study using electrospray ionization mass spectrometry and fluorescence binding studies. *J. Biol. Chem.* 277, 33115–33126.
- Ashkenazy, H., Erez, E., Martz, E., Pupko, T., and Ben-Tal, N. (2010). ConSurf 2010: calculating evolutionary conservation in sequence and structure of proteins and nucleic acids. *Nucleic Acids Res.* 38 (Web Server issue), W529–W533.
- Balch, W.E., Morimoto, R.I., Dillin, A., and Kelly, J.W. (2008). Adapting proteostasis for disease intervention. *Science* 319, 916–919.
- Behrends, C., Langer, C.A., Boteva, R., Böttcher, U.M., Stemp, M.J., Schaffar, G., Rao, B.V., Giese, A., Kretzschmar, H., Siegers, K., and Hartl, F.U. (2006). Chaperonin TrIC promotes the assembly of polyQ expansion proteins into nontoxic oligomers. *Mol. Cell* 23, 887–897.
- Bigotti, M.G., and Clarke, A.R. (2008). Chaperonins: The hunt for the Group II mechanism. *Arch. Biochem. Biophys.* 474, 331–339.
- Bogan, A.A., and Thorn, K.S. (1998). Anatomy of hot spots in protein interfaces. *J. Mol. Biol.* 280, 1–9.
- Bouhouche, A., Benomar, A., Bouslam, N., Chkili, T., and Yahyaoui, M. (2006). Mutation in the epsilon subunit of the cytosolic chaperonin-containing t-complex peptide-1 (Cct5) gene causes autosomal recessive mutilating sensory neuropathy with spastic paraplegia. *J. Med. Genet.* 43, 441–443.
- Chen, L., and Sigler, P.B. (1999). The crystal structure of a GroEL/peptide complex: plasticity as a basis for substrate diversity. *Cell* 99, 757–768.
- Chivian, D., and Baker, D. (2006). Homology modeling using parametric alignment ensemble generation with consensus and energy-based model selection. *Nucleic Acids Res.* 34, e112.
- Clackson, T., Ultsch, M.H., Wells, J.A., and de Vos, A.M. (1998). Structural and functional analysis of the 1:1 growth hormone:receptor complex reveals the molecular basis for receptor affinity. *J. Mol. Biol.* 277, 1111–1128.
- Cong, Y., Schröder, G.F., Meyer, A.S., Jakana, J., Ma, B., Dougherty, M.T., Schmid, M.F., Reissmann, S., Levitt, M., Ludtke, S.L., et al. (2012). Symmetry-free cryo-EM structures of the chaperonin TrIC along its ATPase-driven conformational cycle. *EMBO J.* 31, 720–730.
- Davis, I.W., and Baker, D. (2009). RosettaLigand docking with full ligand and receptor flexibility. *J. Mol. Biol.* 385, 381–392.
- Douglas, N.R., Reissmann, S., Zhang, J., Chen, B., Jakana, J., Kumar, R., Chiu, W., and Frydman, J. (2011). Dual action of ATP hydrolysis couples lid closure to substrate release into the group II chaperonin chamber. *Cell* 144, 240–252.
- Feldman, D.E., Thulasiraman, V., Ferreyra, R.G., and Frydman, J. (1999). Formation of the VHL-elongin BC tumor suppressor complex is mediated by the chaperonin TrIC. *Mol. Cell* 4, 1051–1061.
- Feldman, D.E., Spiess, C., Howard, D.E., and Frydman, J. (2003). Tumorigenic mutations in VHL disrupt folding in vivo by interfering with chaperonin binding. *Mol. Cell* 12, 1213–1224.
- Hartl, F.U., Bracher, A., and Hayer-Hartl, M. (2011). Molecular chaperones in protein folding and proteostasis. *Nature* 475, 324–332.
- Heller, M., John, M., Coles, M., Bosch, G., Baumeister, W., and Kessler, H. (2004). NMR studies on the substrate-binding domains of the thermosome: structural plasticity in the protrusion region. *J. Mol. Biol.* 336, 717–729.
- Hong, S., Choi, G., Park, S., Chung, A.S., Hunter, E., and Rhee, S.S. (2001). Type D retrovirus Gag polyprotein interacts with the cytosolic chaperonin TrIC. *J. Virol.* 75, 2526–2534.
- Hua, Q., Dementieva, I.S., Walsh, M.A., Hallenga, K., Weiss, M.A., and Joachimiak, A. (2001). A thermophilic mini-chaperonin contains a conserved polypeptide-binding surface: combined crystallographic and NMR studies of the GroEL apical domain with implications for substrate interactions. *J. Mol. Biol.* 306, 513–525.
- Hynes, G.M., and Willison, K.R. (2000). Individual subunits of the eukaryotic cytosolic chaperonin mediate interactions with binding sites located on subdomains of beta-actin. *J. Biol. Chem.* 275, 18985–18994.
- Inoue, Y., Aizaki, H., Hara, H., Matsuda, M., Ando, T., Shimoji, T., Murakami, K., Masaki, T., Shoji, I., Homma, S., et al. (2011). Chaperonin TrIC/CCT participates in replication of hepatitis C virus genome via interaction with the viral NS5B protein. *Virology* 410, 38–47.
- Jäger, S., Cimermancic, P., Gulbahce, N., Johnson, J.R., McGovern, K.E., Clarke, S.C., Shales, M., Mercenne, G., Pache, L., Li, K., et al. (2012). Global landscape of HIV-human protein complexes. *Nature* 481, 365–370.

- Kalishman, N., Adams, C.M., and Levitt, M. (2012). Subunit order of eukaryotic TRiC/CCT chaperonin by cross-linking, mass spectrometry, and combinatorial homology modeling. *Proc. Natl. Acad. Sci. USA* **109**, 2884–2889.
- Kasembeli, M., Lau, W.C., Roh, S.H., Eckols, T.K., Frydman, J., Chiu, W., and Tweardy, D.J. (2014). Modulation of STAT3 folding and function by TRiC/CCT chaperonin. *PLoS Biol.* **12**, e1001844.
- Kim, Y.E., Hipp, M.S., Bracher, A., Hayer-Hartl, M., and Hartl, F.U. (2013). Molecular chaperone functions in protein folding and proteostasis. *Annu. Rev. Biochem.* **82**, 323–355.
- Kitamura, A., Kubota, H., Pack, C.G., Matsumoto, G., Hirayama, S., Takahashi, Y., Kimura, H., Kinjo, M., Morimoto, R.I., and Nagata, K. (2006). Cytosolic chaperonin prevents polyglutamine toxicity with altering the aggregation state. *Nat. Cell Biol.* **8**, 1163–1170.
- Lange, O.F., Rossi, P., Sgourakis, N.G., Song, Y., Lee, H.W., Aramini, J.M., Ertekin, A., Xiao, R., Acton, T.B., Montelione, G.T., and Baker, D. (2012). Determination of solution structures of proteins up to 40 kDa using CS-Rosetta with sparse NMR data from deuterated samples. *Proc. Natl. Acad. Sci. USA* **109**, 10873–10878.
- Leitner, A., Joachimiak, L.A., Bracher, A., Mönkemeyer, L., Walzthoeni, T., Chen, B., Pechmann, S., Holmes, S., Cong, Y., Ma, B., et al. (2012). The molecular architecture of the eukaryotic chaperonin TRiC/CCT. *Structure* **20**, 814–825.
- Li, J., and Buchner, J. (2013). Structure, function and regulation of the hsp90 machinery. *Biom. J.* **36**, 106–117.
- Mao, B., Tejero, R., Baker, D., and Montelione, G.T. (2014). Protein NMR structures refined with Rosetta have higher accuracy relative to corresponding X-ray crystal structures. *J. Am. Chem. Soc.* **136**, 1893–1906.
- Melki, R., Batelier, G., Soulié, S., and Williams, R.C., Jr. (1997). Cytoplasmic chaperonin containing TCP-1: structural and functional characterization. *Biochemistry* **36**, 5817–5826.
- Meyer, A.S., Gillespie, J.R., Walther, D., Millet, I.S., Doniach, S., and Frydman, J. (2003). Closing the folding chamber of the eukaryotic chaperonin requires the transition state of ATP hydrolysis. *Cell* **113**, 369–381.
- Pappenberger, G., Wilsher, J.A., Roe, S.M., Counsell, D.J., Willison, K.R., and Pearl, L.H. (2002). Crystal structure of the CCTgamma apical domain: implications for substrate binding to the eukaryotic cytosolic chaperonin. *J. Mol. Biol.* **318**, 1367–1379.
- Pupko, T., Bell, R.E., Mayrose, I., Glaser, F., and Ben-Tal, N. (2002). Rate4Site: an algorithmic tool for the identification of functional regions in proteins by surface mapping of evolutionary determinants within their homologues. *Bioinformatics* **18** (Suppl 1), S71–S77.
- Reissmann, S., Parnot, C., Booth, C.R., Chiu, W., and Frydman, J. (2007). Essential function of the built-in lid in the allosteric regulation of eukaryotic and archaeal chaperonins. *Nat. Struct. Mol. Biol.* **14**, 432–440.
- Reissmann, S., Joachimiak, L.A., Chen, B., Meyer, A.S., Nguyen, A., and Frydman, J. (2012). A gradient of ATP affinities generates an asymmetric power stroke driving the chaperonin TRiC/CCT folding cycle. *Cell Rep.* **2**, 866–877.
- Rinner, O., Seebacher, J., Walzthoeni, T., Mueller, L.N., Beck, M., Schmidt, A., Mueller, M., and Aebersold, R. (2008). Identification of cross-linked peptides from large sequence databases. *Nat. Methods* **5**, 315–318.
- Ritco-Vonsovici, M., and Willison, K.R. (2000). Defining the eukaryotic cytosolic chaperonin-binding sites in human tubulins. *J. Mol. Biol.* **304**, 81–98.
- Rommelaere, H., De Neve, M., Melki, R., Vandekerckhove, J., and Ampe, C. (1999). The cytosolic class II chaperonin CCT recognizes delineated hydrophobic sequences in its target proteins. *Biochemistry* **38**, 3246–3257.
- Rüdiger, S., Buchberger, A., and Bukau, B. (1997). Interaction of Hsp70 chaperones with substrates. *Nat. Struct. Biol.* **4**, 342–349.
- Saibil, H. (2013). Chaperone machines for protein folding, unfolding and disaggregation. *Nat. Rev. Mol. Cell Biol.* **14**, 630–642.
- Salgado, G.F., Marquant, R., Vogel, A., Alves, I.D., Feller, S.E., Morellet, N., and Bouaziz, S. (2009). Structural studies of HIV-1 Gag p6ct and its interaction with Vpr determined by solution nuclear magnetic resonance. *Biochemistry* **48**, 2355–2367.
- Shen, Y., Vernon, R., Baker, D., and Bax, A. (2009). De novo protein structure generation from incomplete chemical shift assignments. *J. Biomol. NMR* **43**, 63–78.
- Spiess, C., Meyer, A.S., Reissmann, S., and Frydman, J. (2004). Mechanism of the eukaryotic chaperonin: protein folding in the chamber of secrets. *Trends Cell Biol.* **14**, 598–604.
- Spiess, C., Miller, E.J., McClellan, A.J., and Frydman, J. (2006). Identification of the TRiC/CCT substrate binding sites uncovers the function of subunit diversity in eukaryotic chaperonins. *Mol. Cell* **24**, 25–37.
- Swain, J.F., Schulz, E.G., and Gierasch, L.M. (2006). Direct comparison of a stable isolated Hsp70 substrate-binding domain in the empty and substrate-bound states. *J. Biol. Chem.* **281**, 1605–1611.
- Tam, S., Geller, R., Spiess, C., and Frydman, J. (2006). The chaperonin TRiC controls polyglutamine aggregation and toxicity through subunit-specific interactions. *Nat. Cell Biol.* **8**, 1155–1162.
- Tam, S., Spiess, C., Auyeung, W., Joachimiak, L., Chen, B., Poirier, M.A., and Frydman, J. (2009). The chaperonin TRiC blocks a huntingtin sequence element that promotes the conformational switch to aggregation. *Nat. Struct. Mol. Biol.* **16**, 1279–1285.
- Tian, G., and Cowan, N.J. (2013). Tubulin-specific chaperones: components of a molecular machine that assembles the α/β heterodimer. *Methods Cell Biol.* **115**, 155–171.
- Trinidad, A.G., Muller, P.A., Cuellar, J., Klejnot, M., Nobis, M., Valpuesta, J.M., and Voudsen, K.H. (2013). Interaction of p53 with the CCT complex promotes protein folding and wild-type p53 activity. *Mol. Cell* **50**, 805–817.
- van der Schot, G., Zhang, Z., Vernon, R., Shen, Y., Vranken, W.F., Baker, D., Bonvin, A.M., and Lange, O.F. (2013). Improving 3D structure prediction from chemical shift data. *J. Biomol. NMR* **57**, 27–35.
- Yam, A.Y., Xia, Y., Lin, H.T., Burlingame, A., Gerstein, M., and Frydman, J. (2008). Defining the TRiC/CCT interactome links chaperonin function to stabilization of newly made proteins with complex topologies. *Nat. Struct. Mol. Biol.* **15**, 1255–1262.
- You, L., Gillilan, R., and Huffaker, T.C. (2004). Model for the yeast cofactor A-beta-tubulin complex based on computational docking and mutagenesis. *J. Mol. Biol.* **341**, 1343–1354.
- Zhou, H., Xu, M., Huang, Q., Gates, A.T., Zhang, X.D., Castle, J.C., Stec, E., Ferrer, M., Strulovici, B., Hazuda, D.J., and Espeseth, A.S. (2008). Genome-scale RNAi screen for host factors required for HIV replication. *Cell Host Microbe* **4**, 495–504.

A Memory System of Negative Polarity Cues Prevents Replicative Aging

Franz Meitinger,^{1,5} Anton Khmelinskii,² Sandrine Morlot,³ Bahtiyar Kurtulmus,¹ Saravanan Palani,^{1,6} Amparo Andres-Pons,^{1,7} Birgit Hub,⁴ Michael Knop,² Gilles Charvin,³ and Gislene Pereira^{1,*}

¹Molecular Biology of Centrosomes and Cilia, German Cancer Research Center (DKFZ), DKFZ-ZMBH Alliance, Im Neuenheimer Feld 581, Heidelberg 69120, Germany

²Center for Molecular Biology of the University of Heidelberg (ZMBH), DKFZ-ZMBH Alliance, 69120 Heidelberg, Germany

³Institut de Génétique et de Biologie Moléculaire et Cellulaire, 1 Rue Laurent Fries, 67400 Illkirch Cedex, France

⁴German Cancer Research Center (DKFZ), Im Neuenheimer Feld 280, 69120 Heidelberg, Germany

⁵Present address: Ludwig Institute for Cancer Research, 9500 Gilman Drive, CMM East, La Jolla, CA 92093, USA

⁶Present address: Division of Biomedical Cell Biology, Warwick Medical School, University of Warwick, Coventry CV4 7AL, UK

⁷Present address: European Molecular Biology Laboratory, Structural and Computational Biology Unit, Meyerhofstrasse 1, 69117 Heidelberg, Germany

*Correspondence: g.pereira@dkfz.de

<http://dx.doi.org/10.1016/j.cell.2014.10.014>

SUMMARY

Cdc42 is a highly conserved master regulator of cell polarity. Here, we investigated the mechanism by which yeast cells never re-establish polarity at cortical sites (cytokinesis remnants [CRMs]) that have previously supported Cdc42-mediated growth as a paradigm to mechanistically understand how Cdc42-inhibitory polarity cues are established. We revealed a two-step mechanism of loading the Cdc42 antagonist Nba1 into CRMs to mark these compartments as refractory for a second round of Cdc42 activation. Our data indicate that Nba1 together with a cortically tethered adaptor protein confers memory of previous polarization events to translate this spatial legacy into a biochemical signal that ensures the local singularity of Cdc42 activation. “Memory loss” mutants that repeatedly use the same polarity site over multiple generations display nuclear segregation defects and a shorter lifespan. Our work thus established CRMs as negative polarity cues that prevent Cdc42 reactivation to sustain the fitness of replicating cells.

INTRODUCTION

The establishment of cell polarity sites is fundamental for a plethora of cellular functions related to morphogenesis, differentiation, and/or proliferation of uni- and multicellular organisms (Bloch and Yalovsky, 2013; Dworkin, 2009; Heasman and Ridley, 2008; Howell and Lew, 2012; Iden and Collard, 2008; Li and Bowerman, 2010; Martin-Belmonte and Perez-Moreno, 2011; McCaffrey and Macara, 2009; Nelson, 2009; Noatynska et al., 2013). The Rho-GTPase Cdc42, initially described in yeast (Johnson and Pringle, 1990), is a master regulator of cell polarization and highly conserved among eukaryotes (Boureux et al., 2007; Etienne-Manneville, 2004). How Cdc42 polarity sites are

regulated in space and time has been extensively studied over the past years (Casamayor and Snyder, 2002; Johnson et al., 2011; Park and Bi, 2007). However, much less is known about the molecular mechanisms involved in the establishment of polarity cue refractory to Cdc42 activation.

During G1 phase, yeast cells establish a polarity site from which the daughter cell will emerge. Polarity establishment depends on Rho GTPase Cdc42 (Park and Bi, 2007). The site of bud emergence (bud neck) will later be used in cytokinesis to separate the daughter from mother cell (Figure 1A) (Park and Bi, 2007; Wloka and Bi, 2012). The remnants of the cytokinetic machinery (bud scar or cytokinesis remnant [CRM]) are retained in the mother cell and can be visualized by transmission electron microscopy (TEM) or by specific dyes (Figure 1B) (Meitinger et al., 2013). CRMs mainly consist of extracellular matrix, which is encircled by a chitin-rich ring (Cabib et al., 1993). In addition, transmembrane proteins, including Rax1 and Rax2, protrude into the extra- and intracellular spaces (Kang et al., 2004). This indicates that CRMs may modulate intracellular processes. However, relatively little is known about the composition and function of CRMs. During the replicative lifetime, the aging mother cell accumulates increasing numbers of these CRMs over its cell surface (Casamayor and Snyder, 2002). Importantly, a new polarity site is never established within CRMs, indicating that Cdc42 activation cannot occur twice at the same site, even though the initiation event could be many generations after the cytokinesis that generated the remnant.

The molecular basis for the prevention of Cdc42 activation at CRMs remains to be established. Different GTPase-activating proteins (GAPs) contribute to the inhibition of Cdc42 at the cell-division site (Atkins et al., 2013; Tong et al., 2007). In addition, the scaffold protein Gps1 inhibits Cdc42 at the site of cytokinesis as part of a pathway working in parallel to the Cdc42 GAP Rga1 (Meitinger et al., 2013). However, neither Cdc42 GAPs nor Gps1 accumulate at old CRMs. Using a combination of proteomics, cell biology, and biochemical approaches, we identified a protein complex that is recruited to the cell-division site by Gps1 and inherited to CRMs in a Gps1 and Rax1-Rax2-dependent manner. We show that Nba1, a bud-neck-associated protein of unknown

function (Calvert et al., 2008), is the core component of this complex that is responsible for the prevention of Cdc42 activation in the remnant. Mutant cells that are unable to inhibit Cdc42 activation at CRMs display nuclear segregation defects and have markedly shorter lifespans. This study therefore identifies the molecular mechanism for a long-recognized phenomenon that keeps CRMs inactive for Cdc42-reactivation, thereby ensuring the longevity of asymmetrically dividing yeast cells.

RESULTS

Nap1, Nba1, and Nis1 Function with Gps1 in Cdc42 Inhibition

Nap1, Nba1, and Nis1 were previously identified as putative Gps1-interacting proteins (Meitinger et al., 2013). Nap1 is a conserved histone chaperone involved in chromatin assembly and morphogenesis (Ishimi and Kikuchi, 1991; Mortensen et al., 2002), whereas Nba1 and Nis1 are proteins of unknown function associated with the cell-division site (Calvert et al., 2008; Iwase and Toh-e, 2001). Using coimmunoprecipitation experiments, we established that Nap1, Nba1, and Nis1 interact with Gps1 (Figure 1C). Fluorescence microscopy analysis revealed that GFP-tagged Gps1, Nap1, Nba1, and Nis1 localize to the cell-division site (Figure 1D). This localization was cell-cycle-dependent for Nap1, Nba1, and Nis1 (Figures S1A and S1B available online). In contrast, only Nba1 and Nis1 localized to CRMs (Figure 1E).

Gps1 localized to the cell-division site independently of *NAP1*, *NBA1*, or *NIS1* (Figures 1F and S1C). Similarly, Nap1 localization to the cell-division site was not affected in *gps1Δ*, *nba1Δ*, or *nis1Δ* cells (Figures 1F and S1C). Nba1 failed to localize to the cell-division site in the absence of *GPS1* (Figures 1F and S1C), although its protein levels were unaffected by deletion of *GPS1* (Figure S1D). Nis1 localization to the cell-division site also required *GPS1*. Moreover, *NAP1* and *NBA1* were necessary for Nis1 localization to the cell-division site (Figures 1F and S1C). Consistent with this observation, coimmunoprecipitation of Nis1 with Gps1 was impaired in the absence of *NAP1* or *NBA1* (Figure 1C).

Interestingly, although GFP-tagged Gps1 and Nap1 did not localize to CRMs, both were necessary for Nba1 and Nis1 localization to CRMs (Figures 1G and S1C). Impaired localization of Nis1 to CRMs in *nap1Δ* cells was at least partially explained by reduced levels of Nis1-GFP in this mutant (Figure S1D). Furthermore, localization of Nba1 and Nis1 to CRMs was mutually dependent (Figure 1G). Together, we conclude that Gps1 and Nap1 are involved in recruiting Nba1 and Nis1 to the cell-division site, and in loading Nba1 and Nis1 to CRMs (Figures 1H).

Gps1 coordinates the activity of Cdc42 and Rho1, two members of the Rho family of small GTPases (Meitinger et al., 2013). In the absence of *GPS1*, abnormal activation of the Cdc42 GTPase causes rebudding inside the old cell-division site, resulting in the appearance of one or more concentric collars of cell wall material at the cell-division site in 30%–40% of the cells (Figure 1I) (Meitinger et al., 2013). In addition, deletion of *GPS1* also compromises the Rho1 cytokinetic pathway, leading to a thinning of the secondary septum (Meitinger et al., 2013). Transmission electron microscopy of *nap1Δ*, *nba1Δ*, and *nis1Δ* cells revealed multiple collars at the bud neck without any thin-

ning of the secondary septum in all three mutants (Table 1A), suggesting that Nap1, Nba1, and Nis1 are involved in regulating Cdc42 but not Rho1. The multiple collar phenotype was more frequent in the *nba1Δ* strain (Table 1A), indicating that Nba1 plays a major role in Cdc42 inhibition. Together, these results established Nap1, Nba1, and Nis1 as Gps1-interacting proteins that are specifically involved in inhibiting the Cdc42 pathway. Furthermore, our findings suggest that inhibition of Cdc42 might not be restricted to the cell-division site but could also occur at CRMs, where Nba1 and Nis1 localize.

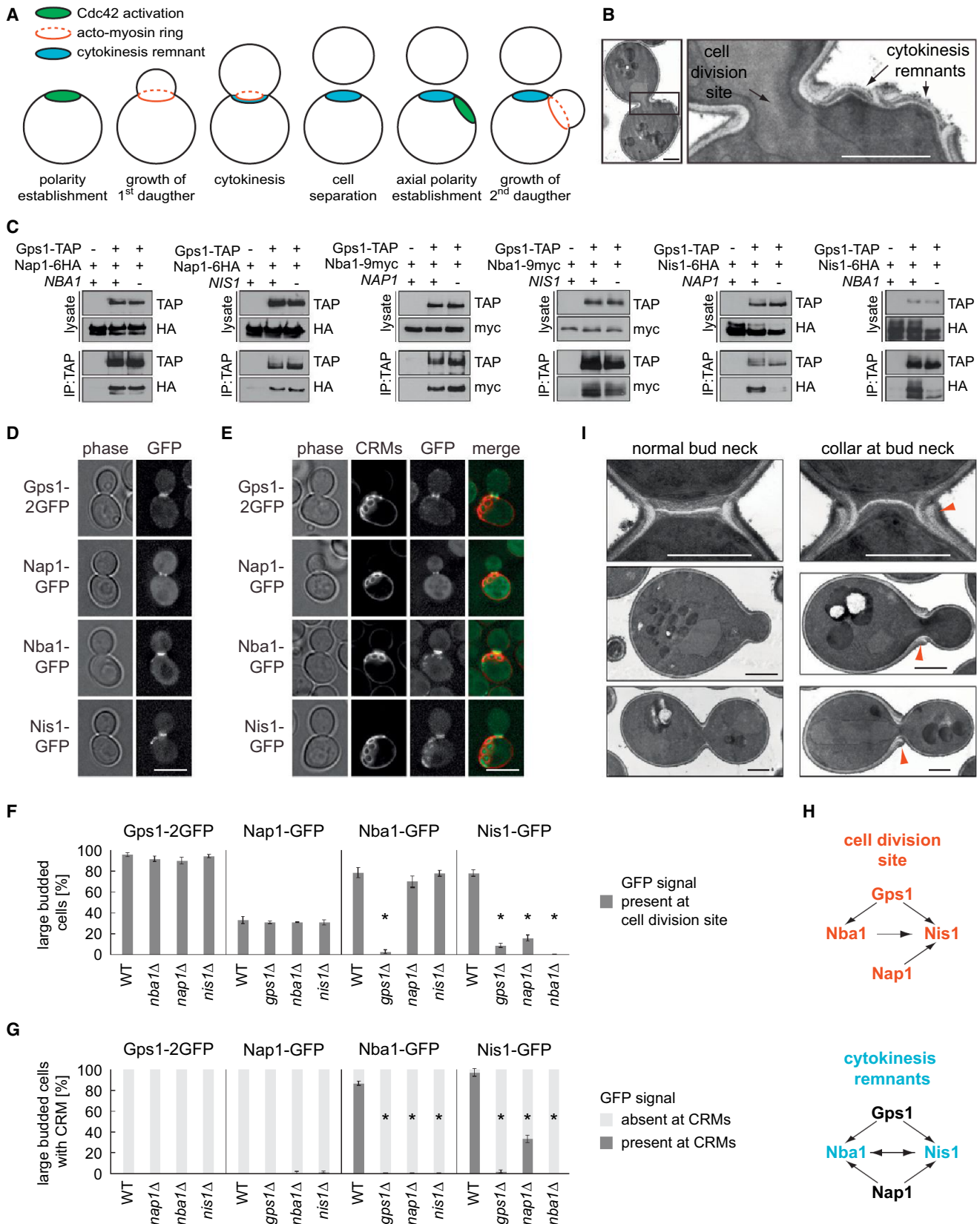
Nba1 Is Necessary and Sufficient to Inhibit Cdc42

We sought to understand how Gps1 interacts with Cdc42, Nap1, Nba1, and Nis1. Yeast two-hybrid experiments with Gps1 truncations suggested that Gps1 has specific binding domains for each of its interaction partners (Figures 2A and S2A). In vitro, recombinant Gps1^{293–422} (a fragment of Gps1 containing the amino acid residues 293–422), Gps1^{514–600} and Gps1^{443–530} interacted directly and specifically with recombinant Nap1, Nba1, and Nis1, respectively (Figure 2B). Based on these results we generated a Gps1^{Δ513–598} mutant in which the Nba1 interaction domain was deleted. Nba1 failed to interact with Gps1^{Δ513–598} in yeast two hybrid (Figure 2C) and showed impaired localization to the cell-division site and CRMs in *gps1^{Δ513–598}* cells (Figures 2D and 2E). Notably, *gps1^{Δ513–598}* cells showed no defects in Rho1 regulation, as judged by the thickness of the secondary septum, yet exhibited multiple collar phenotype with frequency similar to that of *gps1Δ* and *nba1Δ* mutants (Table 1B). This observation suggests that Nba1 functions downstream of Gps1 in the Cdc42 inhibition pathway, and that the main role of Gps1 is to recruit Nba1 to the cell division site and CRMs.

If this model is correct, then artificial targeting of Nba1 to the cell-division site should eliminate the need for Gps1 in Cdc42 inhibition. We tested this possibility using the septin Shs1 tagged with the GFP-binding protein (GBP) (Rothbauer et al., 2008) to recruit GFP fusions to the cell-division site (Meitinger et al., 2013). Expression of *SHS1-GBP* in *NBA1-GFP gps1Δ* cells resulted in permanent association of Nba1-GFP with the cell-division site, partially restored the localization of Nba1-GFP to CRMs and partially rescued the growth defect of the *gps1Δ* mutant (Figures 2F–2H). Notably, artificial tethering of Nba1 to Shs1 rescued the multiple collar phenotype but not the secondary septum defect of *gps1Δ* cells (Table 1C), indicating that *NBA1-GFP gps1Δ SHS1-GBP* cells bypassed the requirement of Gps1 in the regulation of Cdc42 but not of Rho1. Similar results were obtained with *NIS1-GFP gps1Δ SHS1-GBP* cells (Figures S2B–S2D; Table 1D). However, Shs1-tethered Nis1 required *NBA1* for function, whereas Shs1-tethered Nba1 suppressed the *gps1Δ* growth defect even in the absence of *NIS1* (Figures 2H and S2D). Collectively, our data indicate that Gps1 recruits Nba1 and Nis1 to the cell-division site and CRMs, where Nba1 is necessary and sufficient to inhibit Cdc42 activation.

Nba1 Inhibits Binding of the GEF Cdc24 to GTP-Bound Rsr1

Next, we concentrated on the mechanism of Cdc42 inhibition by Nba1. Two parallel mechanisms contribute to the site-directed



(legend on next page)

Table 1. Cdc42 Phenotype and Rho1 Phenotype of Different Mutants

No.	Genotype	Cells with Indicated Number of Collars (%)				Septum Thickness (nm)	p Value	n
		0	1	2	>2	Mean (SD)		
A	1 Wild-type	100	0	0	0	274 (98)	—	108
	2 <i>gps1</i> Δ	64	28	7	1	128 (51)	<0.001	77
	3 <i>nba1</i> Δ	61	31	8	0	231 (85)	0.006	51
	4 <i>nis1</i> Δ	95	5	0	0	266 (98)	0.6	44
	5 <i>nap1</i> Δ	97	3	0	0	361 (107)	<0.001	38
	6 <i>rax1</i> Δ	96	4	0	0	258 (108)	0.4	46
	7 <i>rax2</i> Δ	94	6	0	0	297 (107)	0.2	48
	8 <i>rga1</i> Δ	44	40	14	2	317 (113)	0.01	43
	9 <i>rga1</i> Δ <i>nba1</i> Δ	34	38	22	6	272 (95)	0.9	49
	10 <i>rsr1</i> Δ	100	0	0	0	ND	ND	33
	11 <i>nba1</i> Δ <i>rsr1</i> Δ	100	0	0	0	ND	ND	37
	12 <i>rga1</i> Δ <i>rsr1</i> Δ	81	19	0	0	ND	ND	63
	13 <i>nba1</i> Δ <i>rga1</i> Δ <i>rsr1</i> Δ	72	22	4	2	ND	ND	54
B	1 <i>gps1</i> Δ <i>GPS1</i>	100	0	0	0	318 (140)	—	27
	2 <i>gps1</i> Δ	65	30	5	0	118 (44)	<0.001	37
	3 <i>gps1</i> Δ <i>gps1</i> -Δ301-422	98	3	0	0	274 (98)	0.2	64
	4 <i>gps1</i> Δ <i>gps1</i> -Δ443-530	100	0	0	0	332 (116)	0.6	38
	5 <i>gps1</i> Δ <i>gps1</i> -Δ513-598	62	31	7	0	322 (113)	0.8	42
C	1 <i>NBA1</i> -GFP	100	0	0	0	231 (76)	—	49
	2 <i>NBA1</i> -GFP <i>gps1</i> Δ	60	32	4	4	94 (44)	<0.001	50
	3 <i>NBA1</i> -GFP <i>gps1</i> Δ <i>SHS1</i> -GBP	98	2	0	0	98 (39)	<0.001	44
D	1 <i>NIS1</i> -GFP	100	0	0	0	205 (73)	—	31
	2 <i>NIS1</i> -GFP <i>gps1</i> Δ	62	29	6	3	71 (19)	<0.001	34
	3 <i>NIS1</i> -GFP <i>gps1</i> Δ <i>SHS1</i> -GBP	98	2	0	0	89 (39)	<0.001	42

ND, not determined.

activation of Cdc42, the formation of a trimeric complex composed of the Cdc42 guanine nucleotide exchange factor (GEF) Cdc24, the scaffold protein Bem1 and the p21-activated (PAK) kinase Cla4, which can be reconstituted in vitro using recombinant proteins (Figures S3A and S3B) (Bose et al., 2001; Kozubowski et al., 2008), and the recruitment of Cdc24 by the Ras-related GTPase Rsr1 (Park et al., 1997; Shimada et al., 2004; Smith et al., 2013). We sought to determine whether Nba1 affects activation of Cdc42 by these two mechanisms.

In yeast two-hybrid experiments, Nba1 interacted with Bem1 and Cdc24 (Figure 3A). In vitro binding experiments suggested that these interactions are direct, because recombinant Nba1 associated with recombinant Cdc24 and Bem1 in vitro (Figure 3B). Moreover, Nba1 colocalized with Cdc24 and Bem1 at the cell-division site and with Bem1 at the emerging CRMs (Figures 3C–3F). In contrast, Nis1 did not associate with Cdc24 or Bem1 in the yeast two-hybrid system (Figure 3A) or in vitro (Figure S3C), suggesting that Nis1 does not play a direct role in regulating Cdc24. In vitro competition assays revealed that Cdc24

Figure 1. Nba1, Nis1, and Nap1 Function with Gps1 in the Cdc42 Pathway

(A) The illustration depicts active Cdc42 (green), which leads to the growth of the daughter cell. The actomyosin ring (red) forms at the same site where Cdc42 initially established polarity. After cytokinesis, this site is marked by cytokinesis remnants (blue, bud scars).

(B) Electron micrograph of a dividing yeast cell. The enlarged box shows the cell-division site and two cytokinesis remnants. Scale bars, 1.0 μm.

(C) Immunoprecipitation using cell lysates of the indicated genotypes. TAP, tandem and affinity purification tag.

(D and E) Localization of Gps1, Nap1, Nba1, and Nis1 at the cell-division site (D) and at cytokinesis remnants (CRMs) (E). CRMs were stained with calcofluor. Scale bar, 5 μm.

(F) Quantification of (D). Only large-budded cells were counted, n > 100 cells per strain. *p < 0.0001.

(G) Quantification of (E). Only cells with CRMs were counted, n > 100 cells per strain. *p < 0.0001. See Figure S1C for additional p values of (F) and (G). Data in (F) and (G) are represented as mean ± SEM of three independent experiments.

(H) Diagrams summarizing the dependence in protein localization at the bud neck or CRMs as scored in (F) and (G).

(I) Electron micrographs showing cross-sections of the bud-neck region (cell-division site) of wild-type (normal bud neck) and *nba1*Δ (collar at bud neck) cells. The arrowhead points toward a bud-neck collar, which is indicative of Cdc42 activation at the same site of cytokinesis, thereby causing budding inside the old cell-division site. Scale bar, 1 μm. See also Figure S1.

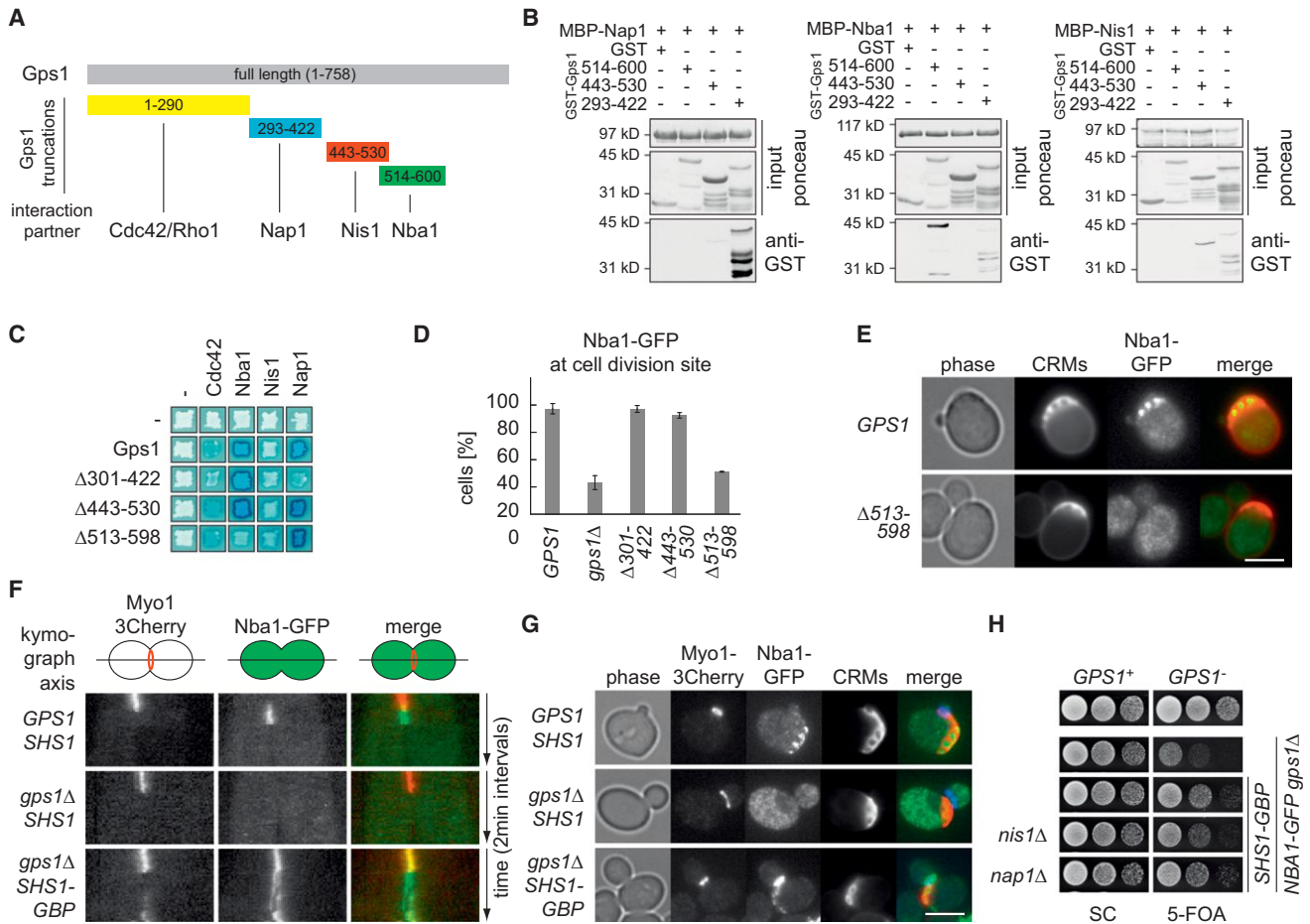


Figure 2. Nba1 Is Required and Sufficient to Inhibit Cdc42 Activation

(A) Schematic representation of the Gps1 fragments that specifically interact with Rho/Cdc42, Nap1, Nis1, and Nba1 (see also Figure S2A). Numbers indicate amino acid positions.

(B) In vitro binding assays showing the binding of GST or the indicated GST-Gps1 fragments to beads coupled to MBP-Nap1, MBP-Nba1, or MBP-Nis1. Ponceau-S-stained nitrocellulose membranes show the inputs for MBP (top panels) and GST (middle panels) fusion proteins.

(C) Yeast two-hybrid analysis of Gps1 full-length and truncated forms lacking the indicated amino acids. Blue color denotes interaction.

(D) Quantification of Nba1 localization at the cell-division site in the indicated Gps1 truncated mutants. Only large-budded cells were counted ($n > 100$ per strain). Data are represented as mean \pm SEM of three independent experiments.

(E) Localization of Nba1 at CRMs. In all inspected cells ($n > 100$ per strain), Nba1 localized at CRMs in *GPS1* but not in *gps1* Δ 513-598 cells. CRMs were stained with calcofluor.

(F) Artificial tethering of Nba1-GFP in wild-type and *gps1* Δ cells at the cell-division site using the GFP binding protein (GBP) fused to the septin Shs1. Kymographs of the daughter-mother axis are shown. Each pixel represents one time point. Myo1-3Cherry, which disappears from the cell-division site after cytokinesis, was used as a cell-cycle marker.

(G) Localization of Nba1-GFP at CRMs in *GPS1*, *gps1* Δ *SHS1*, and *gps1* Δ *SHS1*-GBP cells carrying Myo1-3Cherry. CRMs were stained with calcofluor.

(H) Growth test showing serial dilutions of the strains with the indicated genotypes and carrying an *URA3*-based plasmid expressing wild-type *GPS1*. Cells were spotted on SC-complete plates (maintain *URA3*-*GPS1*) and on 5FOA plates (selects against *URA3*-*GPS1*). Scale bars, 5 μ m. See also Figure S2.

bound to Nba1 only when Bem1 was either absent or present in substoichiometric amounts (Figure 3G, lanes 3–6). Similarly, Bem1 associated with Nba1 only in the presence of low levels of Cdc24 (Figure 3G, lanes 9–12). However, Nba1 was not able to disassemble or to prevent the assembly of Cdc24-Bem1-Cla4 complexes in vitro (Figures S3D and S3E). Therefore, we reasoned that Nba1 binds to Cdc24 and Bem1 to prevent a different activation step.

We next tested whether Nba1 affects the binding of Cdc24 to Rsr1. As previously reported, Cdc24 preferentially bound recom-

binant Rsr1 that was locked in the GTP-bound state (Figure 3H, lanes 1 and 3) (Park et al., 1997). Strikingly, Cdc24 did not associate with GTP-bound Rsr1 in the presence of Nba1 (Figure 3H, lane 2), suggesting that Nba1 could inhibit the activation of Cdc24 by Rsr1. We tested this possibility in vivo using two mutants of Cdc24: Cdc24^{APB}, which cannot bind to Bem1 (Kozubowski et al., 2008), and Cdc24^{G168D}, which is unable to bind Rsr1 (Shimada et al., 2004). Deletion of *NBA1* rescued the growth defect of a *cdc24* Δ strain expressing the *cdc24*^{APB} mutant but not the *cdc24*^{G168D} mutant (Figures 3I and S3F),

suggesting that activation of Cdc24^{APB} by Rsr1 is not sufficient to support growth due to negative regulation by Nba1. In conclusion, our data support a model in which Nba1 locally prevents activation of the guanine nucleotide exchange factor Cdc24 by Rsr1, thereby inhibiting Cdc42 activity at the cell-division site and CRMs (Figure 3J).

The Transmembrane Proteins Rax1 and Rax2 Anchor Nba1 and Nis1 at CRMs

Puzzlingly, Nba1 and Nis1 directly interact with Gps1 and require Gps1 for their localization to both the cell-division site and CRMs, but Gps1 is only associated with the cell-division site. Nba1 and Nis1 have no obvious transmembrane or membrane-binding domains, suggesting they likely rely on a protein other than Gps1 as an anchor at CRMs. Rax1 and Rax2, two transmembrane proteins of unknown function in haploid cells, stably localize to CRMs in an interdependent manner (Chen et al., 2000) (Figures S4A and S4B). We found that *RAX1* and *RAX2* were required for the localization of Nba1 and Nis1 to CRMs but not to the cell-division site, without affecting the total protein levels of Nba1 and Nis1 (Figure 4). Deletion of *RAX1* or *RAX2* impaired Cdc42 inhibition to the same extent as deletion of *NIS1*, as judged by the percentage of cells with multiple collar phenotype (Table 1A). Yeast two-hybrid analysis revealed that Nis1 interacted with the C-terminal cytoplasmic tail of Rax2 in a *NAP1*-dependent manner (Figures S4C–S4E). Deletion of the 37 C-terminal residues of Rax2 (Rax2-ΔC) impaired Nis1 and Nba1 localization to CRMs (Figures S4F and S4G). However, Rax1 also failed to localize to CRMs in *rax2*-ΔC cells (Figure S4H), suggesting that the C-terminal tail of Rax2 is critical for Rax1-Rax2 complex formation. The recruitment of Gps1 or Nap1 to the cell-division site was not influenced by *RAX1* or *RAX2* deletions (data not shown). Together, these results place Rax1 and Rax2 in the Cdc42 inhibition pathway and indicate that Rax1 and Rax2 serve as a binding platform for Nba1 and Nis1 at CRMs.

Nba1 and Nis1 Are Transferred from the Cell-Division Site to the Emerging CRM

We next investigated the mechanism by which Nba1 and Nis1 localize to CRMs. We considered two possibilities: Nba1 and Nis1 could be loaded onto CRMs from a cytoplasmic pool (“diffusion model”) or Nba1 and Nis1 could be transferred, directly or indirectly, from the site of cell division to Rax1-Rax2 at CRMs (“inheritance model”). Two observations suggest that Nba1 and Nis1 must associate with Gps1 and Nap1 at the cell-division site before it becomes a CRM with cell separation, and thus favor the inheritance model. First, Gps1 and Nap1 are restricted to the cell-division site, yet both proteins are required for Nis1 and Nba1 localization to CRMs (Figures 1D–1H). Second, artificial tethering of Nba1 and Nis1 to the cell-division site partially restored Nba1 and Nis1 association with CRMs in *gps1*Δ cells (Figure 2G and S2C). To directly examine Nba1 and Nis1 loading to CRMs, we performed a pulse-chase experiment using cells expressing *NIS1-GFP* under control of the galactose-inducible promoter (*pGal1*) (Figure 5A). Without promoter induction, no Nis1-GFP could be detected in *pGal1-NIS1-GFP* cells (Figure 5A, t = −2 hr). After a 2 hr pulse of induction, Nis1-GFP was detected initially only at the cell-division site

and later at the newly emerging (proximal) but not at the older (distal from the bud neck) CRMs (Figure 5A, t = 0 and 2 hr). Similar results were obtained for Nis1-GFP in pulse-chase experiments with inducible expression of *NBA1* or *RAX1* (Figure S5A). Moreover, Nba1-GFP behaved in an analogous manner in similar experiments (Figures 5A and S5A). These results indicate that Nba1 and Nis1 must associate with the cell-division site before it is converted into a CRM.

Interestingly, Nis1-GFP and Nba1-GFP were found exclusively at distal CRMs 8 hr after the expression pulse (Figure 5A), suggesting that both proteins associate with CRMs in a stable manner. Supporting this notion, analysis of protein age with a tandem fluorescent protein timer (tFT) composed of the fast-maturing green fluorescent protein sfGFP and the slower-maturing red fluorescent protein mCherry (Khmelniskii et al., 2012) showed that Nba1 and Nis1 pools are on average older at CRMs than at the cell-division site (Figure 5B), similar to Rax2, which is known to be stably associated with CRMs (Chen et al., 2000; Khmelniskii et al., 2012). Moreover, in fluorescence recovery after photobleaching experiments, Nis1-GFP showed no detectable turnover at CRMs over a period of more than 4 hr, whereas Nba1-GFP underwent slow turnover at CRMs (half-time >30 min) (Figures 5C, 5D, and S5B; data not shown).

Collectively, our results are consistent with a model in which Nis1 and Nba1 need to associate with the cell-division site to become stably recruited to CRMs, where Nis1 recruits Nba1 through a series of successive steps that required Gps1, Nap1, and Rax1-Rax2 (Figures 1H, 4, and 5E).

Cell-Division Site Reuse Impairs Nuclear Segregation and Shortens Replicative Lifespan

Deletion of the gene coding for Rga1, a GTPase-activating protein (GAP) for Cdc42, was previously reported to result in rebudding at the most recent site of cell division (Tong et al., 2007). We therefore tested whether Rga1 and Nba1 cooperate in preventing the reuse of previous cell-division sites. Transmission electron microscopy showed that the percentage of cells with multiple collars at the bud neck was higher in the *rga1*Δ *nba1*Δ double mutant than in either single mutant (Table 1A). We conclude that Rga1 and Nba1 cooperate in inhibiting rebudding at CRMs. However, whereas Rga1 prevents Cdc42 reactivation at the most recent CRM, Nba1 also prevents Cdc42 reactivation at older CRMs with the help of Gps1, Nap1, Nis1, Rax1, and Rax2.

Next, we sought to determine the physiological significance of preventing rebudding from CRMs. We followed single cells through consecutive divisions until cell death using a microfluidics-based long-term live cell imaging technique (Fehrmann et al., 2013) with a GFP-tagged histone H2B (Htb2) to visualize nuclear segregation (Figure 6A). Whereas wild-type cells divided on average 20 times before death (replicative lifespan) in this assay, the replicative lifespan of cells lacking *NBA1* and *RGF1* alone or in combination was significantly shorter (Figures 6A and 6B). The short replicative lifespan of *nba1*Δ and *rga1*Δ cells was clearly distinct from the frequent premature cell death observed in *gps1*Δ cells (Meitinger et al., 2013). The thinning of the secondary septum observed specifically in *gps1*Δ cells (Table 1A) leads to the lysis of daughter cells during cytokinesis

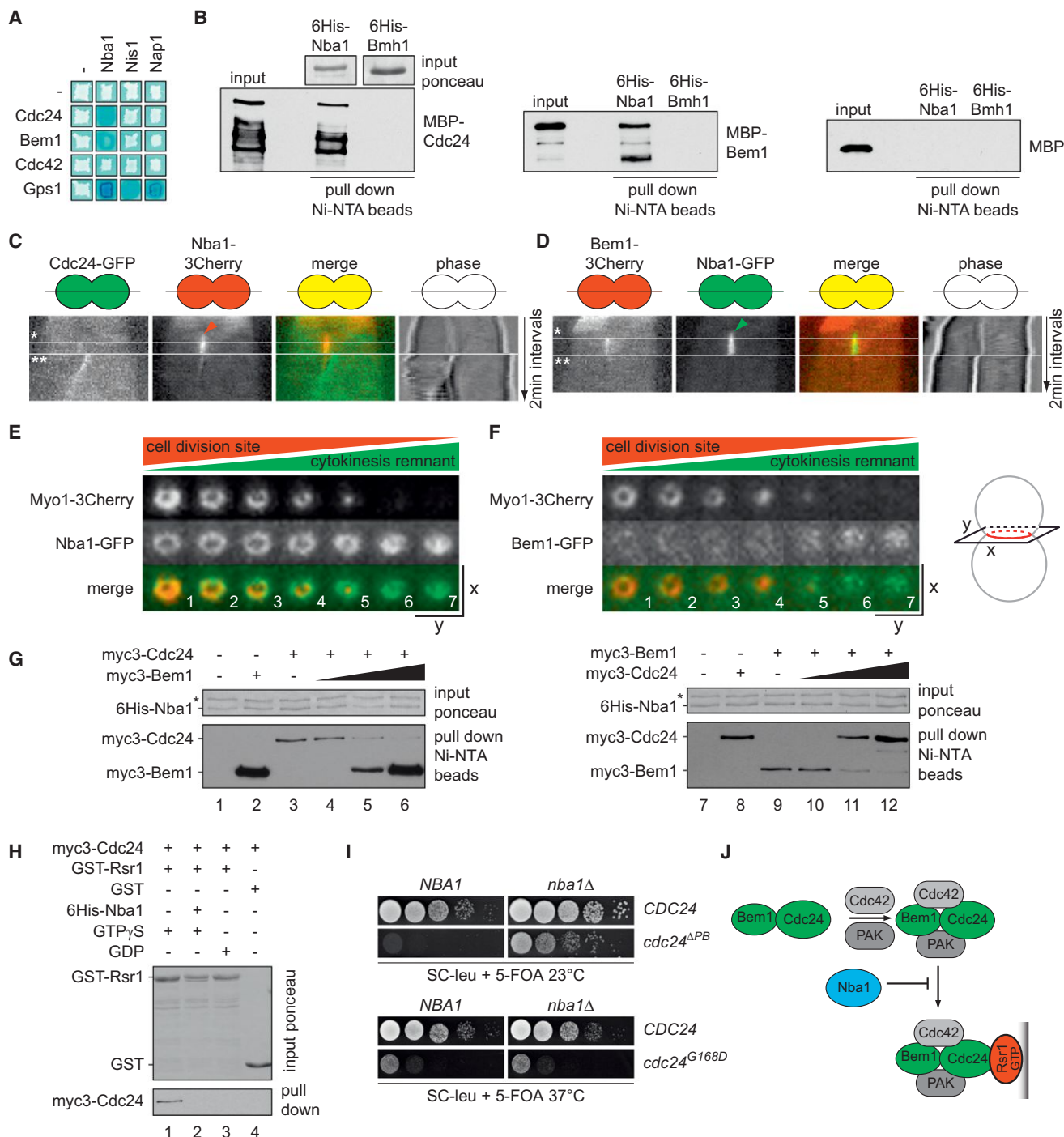


Figure 3. Nba1 Directly Interacts with Cdc24 and Bem1 and Inhibits the Binding of Cdc24 to GTP-Bound Rsr1

(A) Yeast two-hybrid interaction of Nba1, Nis1, and Nap1 with Cdc24, Bem1, Cdc42, and Gps1. The blue color indicates interaction.

(B) In vitro binding assay using recombinant proteins. Immobilized 6His-Nba1 was incubated with MBP alone, MBP-Cdc24, or MBP-Bem1 as indicated. The 14-3-3 family protein Bmh1 was used as a negative control.

(C and D) Kymographs of the daughter-mother cell axis show the colocalization of Cdc24 and Nba1 (C) or Bem1 and Nba1 (D) at the cell-division site. Each pixel represents one time point. Nba1 appears at the cell-division site before Cdc24 (7.1 ± 2.0 min, $n = 15$) or Bem1 (8.1 ± 1.7 min, $n = 17$). Arrowhead, appearance of Nba1 at the cell-division site; *, appearance of Cdc24 and Bem1 at the cell-division site; **, cell separation.

(E and F) Time-lapse series showing top view of Nba1-GFP (E) or Bem1-GFP (F) at the cell-division site during actomyosin ring contraction (Myo1-3Cherry).

(legend continued on next page)

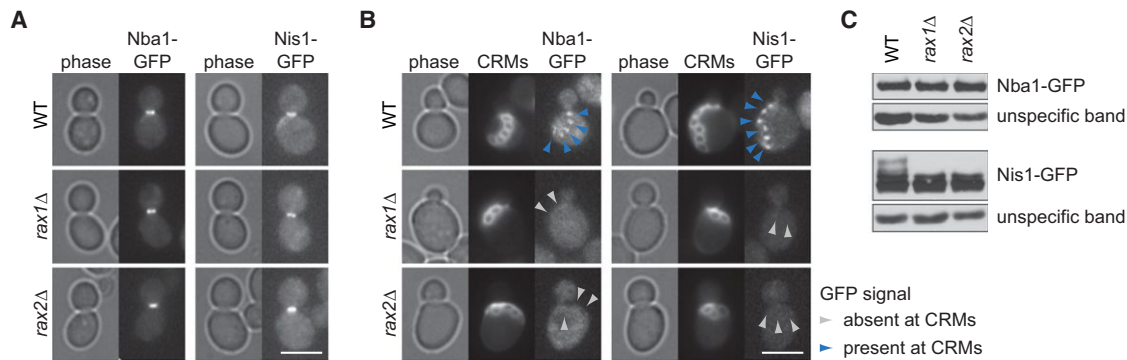


Figure 4. The Transmembrane Proteins Rax1 and Rax2 Target Nba1 and Nis1 to Cytokinesis Remnants

(A and B) Localization of Nba1-GFP and Nis1-GFP at the cell-division site (A) or CRMs (B) in wild-type, *rax1Δ*, and *rax2Δ* cells ($n > 100$ for each strain). Arrows point to CRMs. CRMs were stained with calcofluor. Scale bar, 5 μ m.

(C) Immunoblots showing the protein amounts of Nba1-GFP and Nis1-GFP in wild-type, *rax1Δ*, and *rax2Δ* cells using anti-GFP antibodies. An unspecific band served as a loading control. See also Figure S4.

(Meitinger et al., 2013). As a result, an asynchronous *gps1Δ* (but not *nba1Δ*, *rga1Δ*, or *nba1Δ rga1Δ*) population contained ~30% of dead cells and exhibited reduced doubling time (Figures S6A and S6B).

We thus examined the premature cell death phenotype of the *nba1Δ rga1Δ* mutant in more detail. We determined the time that it took to segregate the nucleus into the daughter cell body in each cell division and complete anaphase throughout the course of replicative aging. Wild-type cells completed nuclear segregation in less than 20 min, with some sporadic delays (Figure 6C). Nuclear segregation delays became more frequent in wild-type cells after entry into senescence, which was scored as a sudden increase in the duration of cell cycle (Fehrmann et al., 2013) (Figures 6C and 6D). In contrast, delayed nuclear segregation was frequent even in young (i.e., nonsenescent) *nba1Δ rga1Δ* cells and often immediately preceded cell death (Figures 6A, 6C, and 6D), suggesting that these cells died prematurely as a consequence of a defective mitosis. Consistently, nuclear segregation into the daughter compartment frequently failed in *nba1Δ rga1Δ* cells, resulting in binucleated cells (Figures 6C and 6D).

How could reuse of previous cell-division sites cause defects in nuclear segregation? We observed that the diameter of the bud-neck opening was narrower in cells with multiple collars, which underwent rebudding from CRMs (Figures 6E and 6F). Moreover, the replicative lifespan of the *nba1Δ*, *rga1Δ*, and *nba1Δ rga1Δ* cells negatively correlated with the frequency of rebudding from CRMs, as judged by the percentage of cells

with bud-neck collars (Figure 6G). We reasoned that narrowing of the bud neck due to repeated rebudding from the same site could hinder nuclear segregation into the daughter cell, thus explaining the nuclear segregation defects and shortened lifespan of *nba1Δ rga1Δ* cells. Supporting this hypothesis, deletion of *RSR1*, which results in Cdc24 mislocalization and random budding (Chant and Herskowitz, 1991), reduced the percentage of *nba1Δ rga1Δ* cells with bud-neck collars (Table 1A) and significantly rescued the survival rate of *nba1Δ rga1Δ* cells (Figure 6H).

Collectively, our data strongly suggest that the relocation of the site of bud growth makes a significant contribution to the fidelity of nuclear segregation and is required to ensure a robust replicative lifespan.

DISCUSSION

Although it is long recognized that Cdc42 activation is never established twice at the same site in yeast, the underlying molecular mechanisms and physiological significance of such a regulation remain elusive. Here, we show that CRMs of budding yeast are marked by Cdc42 inhibitory complexes. We identified the protein Nba1 as a core subunit of these complexes and show that Nba1 directly regulates Cdc42. In vitro analyses combined with in vivo functional studies establish that Nba1 spatially inhibits the activation of Cdc42 by impeding the interaction between the Cdc24 GEF, Cdc42, and its upstream activator, the GTPase Rsr1. This study therefore reveals Nba1 as a player that

(G) In vitro competition assay using bacterial purified proteins. Left panel: immobilized Nba1 was incubated with buffer only (lane 1), myc3-Bem1 (lane 2), and myc3-Cdc24 in the absence (lane 3) or presence of increasing amounts of Bem1 (lanes 4–6). Right panel: same as before but using a constant amount of myc3-Bem1 and variable amounts of myc3-Cdc24. Asterisk, unspecific protein.

(H) Nba1 interrupts the interaction of GTP-Rsr1 with Cdc24. Immobilized GST-Rsr1 was incubated with GTP γ S (lanes 1 and 2), GDP (lane 3), or buffer only (lane 4). GST-Rsr1 beads were incubated with myc3-Cdc24 protein extract in the absence (lane 1) or presence (lane 2) of 6His-Nba1. GST only was used as a negative control (lane 4). Myc3-Cdc24 was detected using anti-myc antibodies.

(I) Growth of *cdc24Δ NBA1 URA3-CDC24* or *cdc24Δ nba1Δ URA3-CDC24* cells carrying a *LEU2*-based plasmid containing full-length *CDC24*, *cdc24^{ΔPB}* (*CDC24* codons 1–750), or *CDC24^{G168D}* (glycine 168 substituted to glutamic acid) as indicated. Serial dilution of cells were spotted on SC-LEU (*URA3-CDC24* is retained) and 5FOA-LEU2 (*URA3-CDC24* is lost) plates and incubated at 23°C or 37°C for 3 days (see Figure S4E for growth at different temperatures).

(J) Summary of Nba1 interactions. Our data indicate that Nba1 binds directly to Bem1 and Cdc24 and inhibits Cdc24 binding to Rsr1 at CRMs. See also Figure S3.

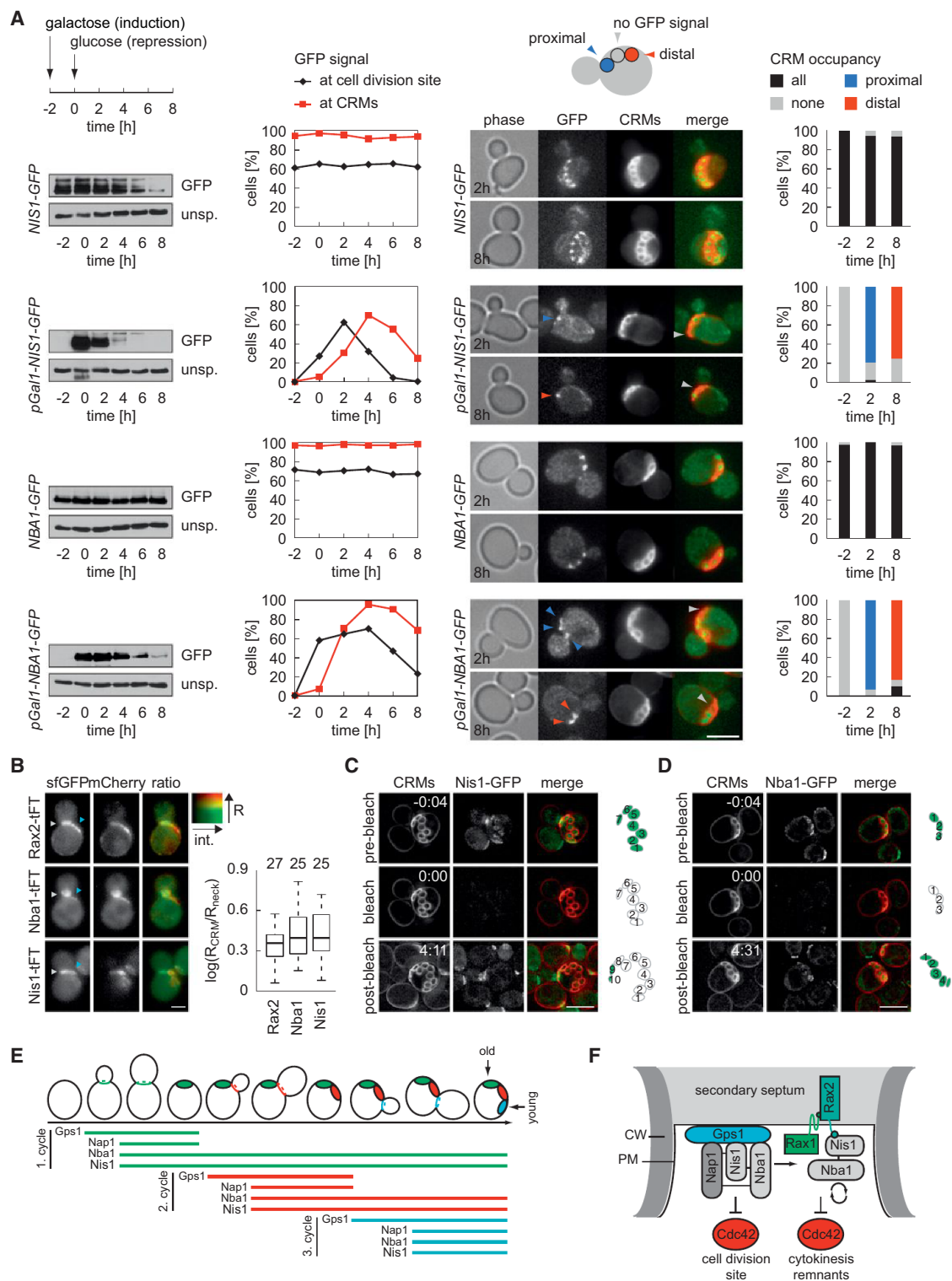


Figure 5. Nba1 Is Inherited at Cytokinesis Remnants by a Two-Step Mechanism

(A) Pulse-chase experiments using *NIS1-GFP*, *pGal1-NIS1-GFP*, *NBA1-GFP*, or *pGal1-NBA1-GFP* cells were performed as illustrated in the time diagram. The expression of genes under control of the galactose promoter was induced for a period of 2 hr, allowing the accumulation of the corresponding protein. After repression of the galactose promoter (addition of glucose, $t = 0$), the localization of the indicated GFP-tagged proteins was analyzed over time. CRMs were stained with calcofluor. Newly formed CRMs located proximal to the new emerging daughter cell (blue), whereas older ones locate more distally (red). Protein

(legend continued on next page)

constrains Rsr1 function and acts alongside landmark proteins, which are involved in the spatial control of bud site selection (Casamayor and Snyder, 2002). Our data further establish that the inhibitory mode of Nba1 action upon Cdc42 differs from the control of GTP hydrolysis that is orchestrated by the GAP protein Rga1. Accordingly, epistasis studies established that Rga1 and Nba1 collaborate in the inhibition of Cdc42, because cells lacking both proteins reused the same polarity site for Cdc42 activation with a higher frequency than cells lacking either regulator alone. However, in contrast to Rga1, which is confined to the site of cell division, Nba1 remains associated with older CRMs. Our data thus indicate that Nba1 functions as a molecular spatial memory that confers protection from recurrent Cdc42 polarization at the same site. We propose that CRMs are marked with “negative polarity cues” that prevent Cdc42 reactivation.

We previously reported that Gps1 binds Cdc42 and prevents Cla4-mediated polarity establishment at the cell-division site (Meitinger et al., 2013). However, it was not clear how this inhibition was achieved on a molecular level. We found that Gps1 directly interacts with Nba1 and Cdc42. Because Gps1 is dispensable for the inhibition of Cdc42-Cla4-mediated polarity establishment upon artificial targeting of Nba1 to the cell-division site (Figure 2), the inhibition of Cla4 by Gps1 is most likely a consequence of the upstream inhibition of Cdc42 that occurs at the level of Cdc24. We thus favor the idea that Gps1 acts as a scaffold protein at the cell-division site that supports local inactivation of Cdc42 by Nba1.

We describe an elaborated system of three successive steps (recruitment, transfer/inheritance, and maintenance) that are absolutely required to load Nba1 onto CRMs. First, Gps1 recruits Nba1 and the adaptor protein Nis1 to the cell-division site. At this site, the chaperone Nap1 is essential for the transfer of Nba1 and Nis1 from Gps1 to the transmembrane proteins Rax1-Rax2. Nba1 then remains associated with Rax2 at CRMs in a Nis1-dependent manner. Based on localization and deletion analysis, we propose the existence of two complexes that associate with CRMs of different ages. The Gps1/Nba1/Nis1/Nap1 complex is present only at emerging CRMs (cell-division site), whereas the Rax1/Rax2/Nis1/Nba1 complex associates with older CRMs and maintains this association over multiple generations. Importantly, disruption of Nba1 at emergent CRMs causes a more

severe Cdc42 inhibition phenotype than Nba1 removal from established older remnants (e.g., in *nis1Δ*, *nap1Δ*, *rax1Δ*, or *rax2Δ* cells). We therefore reasoned that inhibition of Cdc42 at the most recent site of cell division is more critical than its inhibition at older CRMs. This might explain why cells also rely on the GAP Rga1 to prevent Cdc42 activation at the cell-division site (Tong et al., 2007).

Why have cells developed such an elaborated mechanism to maintain and protect CRMs? In the absence of Nba1 and/or Rga1, the replicative lifespan of yeast cells was significantly shorter than in wild-type controls. This shortening of lifespan correlated with nuclear segregation defects that, in most cases, immediately preceded cell death. We previously reported that cells lacking *GPS1* had an increased rate of cell death that was exacerbated by codepletion of *RGA1* (Meitinger et al., 2013). However, several evidences support that cell death of *nba1Δ rga1Δ* and *gps1Δ* cells are caused by different mechanisms. First, *nba1Δ rga1Δ* cells do not show a defect in cell growth as *gps1Δ* cells do (Figure S6). Second, in contrast to *gps1Δ* mutants, whose daughter cells selectively die during cytokinesis (Meitinger et al., 2013), we mainly observed cell death of large-budded *nba1Δ rga1Δ* cells shortly after binucleation (Figure 6). Third, the growth defect of *gps1Δ* cells is based on the misregulation of the Rho1-pathway (Meitinger et al., 2013), which is functional in *nba1Δ rga1Δ* cells (Table 1). We thus propose that the narrowing of the bud-neck diameter by repeated reuse of the same polarity site was the most likely cause of nuclear segregation defects and cell death in *nba1Δ rga1Δ* cells. In support of this assumption, we found that deletion of *RSR1*, which decreased the probability of reuse of the cell-division site for bud growth, improved the survival rate of *nba1Δ rga1Δ* cells. We thus propose that the combined impact of the different Cdc42 cell polarity control mechanisms guarantees genome stability and the fitness of dividing cells. In light of our study, many open questions remain. For example, it is unclear why yeast cells developed such a complex system for protecting CRMs against Cdc42 repolarization when other cell types simply remove CRMs (Pohl and Jentsch, 2009). It also begs the question as to whether CRMs may play other mother cell-specific functions that are yet to be appreciated, perhaps during the aging process itself.

levels were analyzed using anti-GFP antibodies. An unspecific band served as a loading control. The graphs represent the quantification of the percentage of large-budded cells showing Nba1-GFP at the cell-division site or CRMs (only cells with CRMs were counted). Images show representative cells analyzed at early (2 hr) or late (8 hr) time points after repression of the galactose promoter. The graphs on the right indicate the percentage of cells with the given GFP-tagged protein at none, all, only proximal, or only distal CRMs (“CRM occupancy”). Only cells with more than one CRM were counted. $n > 100$ cells per time point. CRMs were stained with calcofluor.

(B) Nba1, Nis1, and Rax2 are fused to the tandem fluorophore sfGFP-mCherry. The ratio between mCherry and sfGFP fluorescence intensities at the cytokinesis remnants (R_{CRM}) and the cell-division site (R_{neck}) are used to determine the age of the different protein pools. The graph shows the logarithm of the ratio of R_{CRM} and R_{neck} . $\log(R_{CRM}/R_{neck}) > 0$ indicates that the protein pool at cytokinesis remnants is older than at the cell-division site.

(C and D) Fluorescence recovery of Nis1-GFP (C) and Nba1-GFP (D) at CRMs 4.5 hr after photobleaching. The numbers on the right indicate CRMs pre- and postbleach; the green color indicates the presence of the GFP fusion protein at the given CRM. CRMs were stained with calcofluor.

(E) Model for the timely loading of Nba1 and Nis1 to CRMs based on pulse-chase (A), age determination (B), FRAP experiments (C and D), and localization studies (Figure 1). The aforementioned data indicate that Gps1 is recruited to the newly formed bud neck and recruits Nba1 and Nis1 to the cell-division site of large-budded cells. After cell separation, Nis1 and Nba1 but not Gps1 or Nap1 remain stably associated with the emerging CRM as the new bud emerges. Each color represents one generation in which the proteins are attached to CRMs.

(F) Model for the formation of a Gps1 dependent complex at the cell-division site. During cytokinesis, Nba1 and Nis1 are transferred to the transmembrane protein Rax2, which depends on the chaperone Nap1. Both complexes are necessary to prevent Cdc42 activation at the cell-division site and at CRMs. CW, cell wall; PM, plasma membrane and secondary septum are indicated. Scale bars, 5 μ m. See also Figure S5.

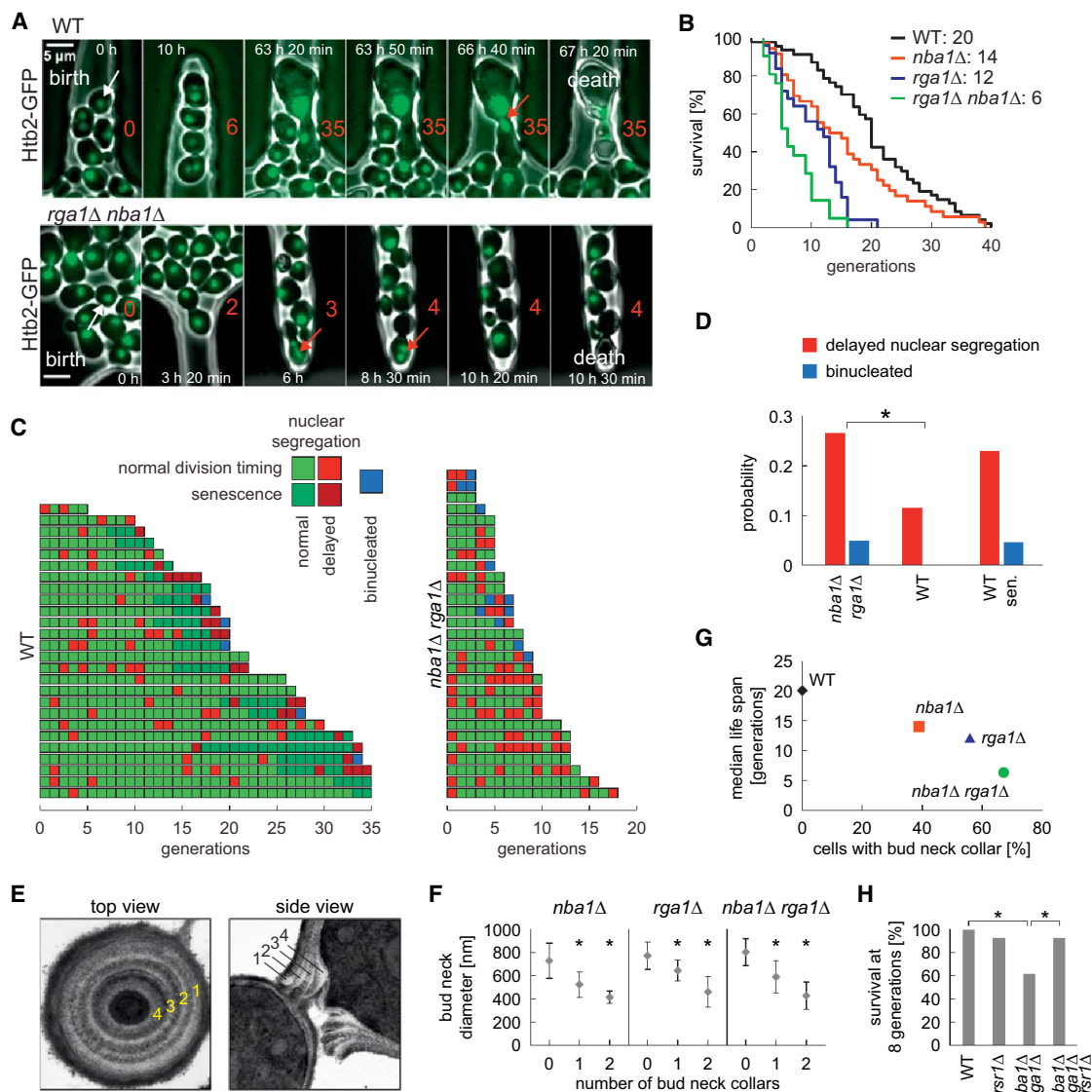


Figure 6. Inhibition of Cdc42 at Cytokinesis Remnants Contributes to the Fidelity of Nuclear Segregation and Longevity

(A) Time-lapse series of wild-type (WT) and *nba1Δ rga1Δ* single cells from birth to death in the microfluidic device at the indicated time points. The histone, Htb2 tagged with GFP, was used as a nuclear marker. The white arrows indicate the cells that became trapped in the cavity. Red numbers indicate how many buds the cell generated since birth. Red arrows mark nuclear segregation delays and binucleated cells.

(B) Survival probability curves of wild-type (WT, $n = 47$), *nba1Δ* ($n = 36$), *rga1Δ* ($n = 25$), and *nba1Δ rga1Δ* ($n = 21$) cells. The median survival is indicated. All mutants display a significantly reduced longevity compared to WT ($p < 0.001$).

(C) Temporal trajectories showing successive divisions of single cells. Green, red, and blue colors indicate normal anaphase (< 20 min), delayed anaphase (> 20 min), and binucleated cells, respectively. Dark colors indicate when cells have entered senescence.

(D) Quantification of the probability per generation of delayed nuclear segregation (> 20 min, red bars) or binucleation (blue bars) to occur. The difference between *nba1Δ rga1Δ* and young wild-type cells was significant ($p < 0.001$). Senescent wild-type cells were scored separately.

(E) TEM transverse (top view) and longitudinal (side view) sections of *nba1Δ* cells with four collars at the bud-neck region, as indicative of recurrent rebudding at the same location.

(F) Quantification of the bud-neck diameter of *nba1Δ*, *rga1Δ*, and *nba1Δ rga1Δ* cells with different numbers of bud-neck collars (no collar, $n = 20$; one collar, $n = 18$; two collar, $n = 18$). Data are represented as mean \pm SEM. $*p < 0.001$.

(G) Graph depicting the percentage of cells with multiple collars (determined in F) versus the median lifespan (determined in B) for the indicated strains.

(H) Percentage of surviving cells at eight generations for the indicated cell types was calculated based on single-cell analysis performed as described in (A) ($n = 40$ for each strain). Note that, due to the random budding pattern of *rsr1Δ* cells, the time that the mother cell remained in the inspection chamber was limited to seven to nine generations. $*p < 0.001$. See also Figure S6.

Molecular remnants of the cell-division site persist at the plasma membrane of daughter cells in several organisms and do so in a cell-type-specific manner. For example, stem cells mimic budding yeast, in accumulating division remnants in successive cell divisions (Chen et al., 2013). Thus, it is tempting to speculate that also in other organisms the cytokinetic machinery may act as platform that allows the specific and asymmetric inheritance of associated protein complexes, which might affect cell fate determination. CRMs have recently been implicated in determining the spatial orientation and establishment of cell polarity in neurons (Pollarolo et al., 2011; Wilcock et al., 2007). Interestingly, Cyk4 (a metazoan inhibitor of the Rho GTPase Rac1) persists at cell-division remnants of *C. elegans* (Bastos et al., 2012; Jantsch-Plunger et al., 2000), and this persistence has been implicated in positioning the centrosome and cell-division plane during early embryonic cell divisions (Hyman, 1989; Keating and White, 1998; Waddle et al., 1994). Thus, it is tempting to speculate that CRMs play a conserved role in affecting cell polarity by acting as a spatial memory in a variety of organisms ranging from yeast to higher eukaryotes.

EXPERIMENTAL PROCEDURES

Yeast Methods and Plasmid Construction

Yeast strains and plasmids used in this study are listed in Tables S1 and S2. For bacterial expression, genes or gene fragments were cloned into pGEX-5X-1 (GST, GE Healthcare), pET28c (6His, Novagen), and pMAL-c2x (MBP, NEB). Yeast growth and media were as described (Sherman, 1991). Gene deletions and epitope tagging were performed using PCR-based methods (Janke et al., 2004; Knop et al., 1999). Genes of interest were expressed from their endogenous promoter unless specified otherwise. Yeast strains were grown in yeast peptone dextrose medium containing 0.1 mg/l adenine (YPAD). Strains carrying plasmids were grown in synthetic complete (SC) media lacking the corresponding amino acids. Loss of *URA3*-containing plasmids was assessed by using plates containing 1 mg/ml 5-fluoroorotic acid (5FOA).

Genetic Interactions Based on Growth

The plasmid shuffle strategy was used to test the viability of single and double mutants (genetic interaction). Briefly, mutant strains containing the corresponding wild-type gene on an *URA3*-based plasmid (pRS316) were analyzed for growth on 5FOA plates (selection against *URA3*). At least six individual transformants were analyzed for each mutant, and one representative mutant is shown. For each mutant 10-fold serial dilutions of wild-type (ESM356-1) and mutants with the indicated genotype were spotted onto SC and 5FOA plates. Mutants were complemented by pRS316-*GPS1* (Figures 2H and S2D) or pRS316-*CDC24* (Figures 3I and S3G). Strains carrying genes on a *LEU2*-based plasmid (Figures 2H, 3I, and S3G) were grown on medium lacking leucine to avoid plasmid loss. Plates were incubated for 1 or 2 days at 30°C.

Protein Detection Methods and Quantifications

Yeast protein extracts and Immunoblotting were performed as described (Janke et al., 2004). Antibodies were rabbit anti-GFP antibody, mouse anti-tubulin (Tub1), mouse anti-HA (clone 12CA5, Sigma), mouse anti-Myc (clone 9E10, Sigma), rabbit anti-Cib2, rabbit anti-TAP (Open Biosystems), mouse anti-MBP (NEB), and guinea pig anti-GST. Secondary antibodies were goat anti-mouse, goat anti-rabbit, and goat anti-guinea pig immunoglobulin Gs coupled to horseradish peroxidase (Jackson ImmunoResearch Laboratories). Protein concentrations were determined using Bradford reagent according to manufactures' instructions (Bio-Rad). Immunoprecipitation, in vitro binding, and yeast two-hybrid assays are described in Extended Experimental Procedures.

Microscopy Techniques

For fluorescence still image analysis, cells carrying GFP or Cherry fusion proteins were inspected without fixation (Figures 1D, 1E, 2A–2D, 3C–3F, 5B–5D, S1B, S2B, S4A, S4B, S4F–S4H, S5B, and S6A) or after fixation with 4% formaldehyde for 20 min (Figures 2E, 2G, 4A, 4B, 5A, S2C, and S5A). CRMs were stained with 0.1 mg/ml calcofluor (Fluorescence Brightener 28, Sigma-Aldrich). Live-cell imaging and quantification of fluorescence still images were performed as described in Extended Experimental Procedures. The tandem fluorescent protein timer experiment for Rax2, Nba1, and Nis1 was performed and quantified as described in Extended Experimental Procedures (Khmelninskii et al., 2012). Long-term FRAP analysis for Nba1 and Nis1 was performed in observation chambers of a Y4C microfluidic plate (CellAsic) at 30°C as described in Extended Experimental Procedures (Khmelninskii et al., 2012). Specimens for electron microscopy were prepared as described previously in Extended Experimental Procedures (Meitinger et al., 2011).

Replicative Lifespan Analysis in a Microfluidic Device

Time-lapse images (phase contrast and fluorescence) were acquired using a microfluidic device as described previously (Fehrmann et al., 2013) using an automated epifluorescence microscope (Nikon Eclipse Ti) with a digital CMOS camera (Hamamatsu Orca Flash 4.0). Cells were grown in synthetic complete medium and then were loaded into the device and allowed to divide until death during typically 100 hr. Constant media perfusion allowed keeping the cells in exponential growth during the course of the experiment. Replicative lifespans were analyzed from time-lapse images (one frame every 10 min). Survival curves were calculated from the number of buds produced by individual mother cells. Nucleus segregation defects were scored using a nuclear marker (Htb2-GFP).

SUPPLEMENTAL INFORMATION

Supplemental Information includes Extended Experimental Procedures, six figures, and two tables and can be found with this article online at <http://dx.doi.org/10.1016/j.cell.2014.10.014>.

ACKNOWLEDGMENTS

We thank Daniel Lew for reagents; Karsten Richter and DKFZ electron microscopy facility for support with electron microscopy; Daniela T. Bertazzi Lazzarini for working with Nba1 and Nis1 at the initial phase of this project; Elmar Schiebel, Iain Hagan, and members of G.P.'s lab for comments on the manuscript. G.C. was supported by the Atip Avenir program and the IDEX program from the University of Strasbourg. S.M. acknowledges a fellowship from the French Ligue Nationale contre le Cancer. M.K. and A.K. were supported by the Sonderforschungsbereich 1036 (SFB1036, TP10) from the DFG. The work of F.M., S.P., and A.A.-P. was supported by the Marie Curie Grant MEXT-CT-042544 granted to G.P. S.P. acknowledges funding from the graduate school LGFG "Regulation of Cell Division". S.P. and B.K. are members of the international graduate school HBIGS. This project was funded by the Marie Curie Grant MEXT-CT-042544 and DKFZ core funding.

Received: March 10, 2014

Revised: August 20, 2014

Accepted: September 25, 2014

Published: November 13, 2014

REFERENCES

- Atkins, B.D., Yoshida, S., Saito, K., Wu, C.F., Lew, D.J., and Pellman, D. (2013). Inhibition of Cdc42 during mitotic exit is required for cytokinesis. *J. Cell Biol.* 202, 231–240.
- Bastos, R.N., Penate, X., Bates, M., Hammond, D., and Barr, F.A. (2012). CYK4 inhibits Rac1-dependent PAK1 and ARHGEF7 effector pathways during cytokinesis. *J. Cell Biol.* 198, 865–880.

- Bloch, D., and Yalovsky, S. (2013). Cell polarity signaling. *Curr. Opin. Plant Biol.* **16**, 734–742.
- Bose, I., Irazoqui, J.E., Moskow, J.J., Bardes, E.S., Zyla, T.R., and Lew, D.J. (2001). Assembly of scaffold-mediated complexes containing Cdc42p, the exchange factor Cdc24p, and the effector Cla4p required for cell cycle-regulated phosphorylation of Cdc24p. *J. Biol. Chem.* **276**, 7176–7186.
- Boureux, A., Vignal, E., Faure, S., and Fort, P. (2007). Evolution of the Rho family of ras-like GTPases in eukaryotes. *Mol. Biol. Evol.* **24**, 203–216.
- Cabib, E., Mol, P.C., Shaw, J.A., and Choi, W.J. (1993). Biosynthesis of cell wall and septum during yeast growth. *Arch. Med. Res.* **24**, 301–303.
- Calvert, M.E., Keck, K.M., Ptak, C., Shabanowitz, J., Hunt, D.F., and Pemberton, L.F. (2008). Phosphorylation by casein kinase 2 regulates Nap1 localization and function. *Mol. Cell. Biol.* **28**, 1313–1325.
- Casamayor, A., and Snyder, M. (2002). Bud-site selection and cell polarity in budding yeast. *Curr. Opin. Microbiol.* **5**, 179–186.
- Chant, J., and Herskowitz, I. (1991). Genetic control of bud site selection in yeast by a set of gene products that constitute a morphogenetic pathway. *Cell* **65**, 1203–1212.
- Chen, T., Hiroko, T., Chaudhuri, A., Inose, F., Lord, M., Tanaka, S., Chant, J., and Fujita, A. (2000). Multigenerational cortical inheritance of the Rax2 protein in orienting polarity and division in yeast. *Science* **290**, 1975–1978.
- Chen, C.T., Ettinger, A.W., Huttner, W.B., and Doxsey, S.J. (2013). Resurrecting remnants: the lives of post-mitotic midbodies. *Trends Cell Biol.* **23**, 118–128.
- Dworkin, J. (2009). Cellular polarity in prokaryotic organisms. *Cold Spring Harb. Perspect. Biol.* **1**, a003368.
- Etienne-Manneville, S. (2004). Cdc42—the centre of polarity. *J. Cell Sci.* **117**, 1291–1300.
- Fehrmann, S., Paoletti, C., Goulev, Y., Ungureanu, A., Aguilaniu, H., and Charvin, G. (2013). Aging yeast cells undergo a sharp entry into senescence unrelated to the loss of mitochondrial membrane potential. *Cell Reports* **5**, 1589–1599.
- Heasman, S.J., and Ridley, A.J. (2008). Mammalian Rho GTPases: new insights into their functions from in vivo studies. *Nat. Rev. Mol. Cell Biol.* **9**, 690–701.
- Howell, A.S., and Lew, D.J. (2012). Morphogenesis and the cell cycle. *Genetics* **190**, 51–77.
- Hyman, A.A. (1989). Centrosome movement in the early divisions of *Caenorhabditis elegans*: a cortical site determining centrosome position. *J. Cell Biol.* **109**, 1185–1193.
- Iden, S., and Collard, J.G. (2008). Crosstalk between small GTPases and polarity proteins in cell polarization. *Nat. Rev. Mol. Cell Biol.* **9**, 846–859.
- Ishimi, Y., and Kikuchi, A. (1991). Identification and molecular cloning of yeast homolog of nucleosome assembly protein I which facilitates nucleosome assembly in vitro. *J. Biol. Chem.* **266**, 7025–7029.
- Iwase, M., and Toh-e, A. (2001). Nis1 encoded by YNL078W: a new neck protein of *Saccharomyces cerevisiae*. *Genes Genet. Syst.* **76**, 335–343.
- Janke, C., Magiera, M.M., Rathfelder, N., Taxis, C., Reber, S., Maekawa, H., Moreno-Borchart, A., Doenges, G., Schwob, E., Schiebel, E., and Knop, M. (2004). A versatile toolbox for PCR-based tagging of yeast genes: new fluorescent proteins, more markers and promoter substitution cassettes. *Yeast* **21**, 947–962.
- Jantsch-Plunger, V., Gönczy, P., Romano, A., Schnabel, H., Hamill, D., Schnabel, R., Hyman, A.A., and Glotzer, M. (2000). CYK-4: A Rho family gtpase activating protein (GAP) required for central spindle formation and cytokinesis. *J. Cell Biol.* **149**, 1391–1404.
- Johnson, D.I., and Pringle, J.R. (1990). Molecular characterization of CDC42, a *Saccharomyces cerevisiae* gene involved in the development of cell polarity. *J. Cell Biol.* **111**, 143–152.
- Johnson, J.M., Jin, M., and Lew, D.J. (2011). Symmetry breaking and the establishment of cell polarity in budding yeast. *Curr. Opin. Genet. Dev.* **21**, 740–746.
- Kang, P.J., Angerman, E., Nakashima, K., Pringle, J.R., and Park, H.O. (2004). Interactions among Rax1p, Rax2p, Bud8p, and Bud9p in marking cortical sites for bipolar bud-site selection in yeast. *Mol. Biol. Cell* **15**, 5145–5157.
- Keating, H.H., and White, J.G. (1998). Centrosome dynamics in early embryos of *Caenorhabditis elegans*. *J. Cell Sci.* **111**, 3027–3033.
- Khmelniskii, A., Keller, P.J., Bartosik, A., Meurer, M., Barry, J.D., Mardin, B.R., Kaufmann, A., Trautmann, S., Wachsmuth, M., Pereira, G., et al. (2012). Tandem fluorescent protein timers for in vivo analysis of protein dynamics. *Nat. Biotechnol.* **30**, 708–714.
- Knop, M., Siegers, K., Pereira, G., Zachariae, W., Winsor, B., Nasmyth, K., and Schiebel, E. (1999). Epitope tagging of yeast genes using a PCR-based strategy: more tags and improved practical routines. *Yeast* **15** (10B), 963–972.
- Kozubowski, L., Saito, K., Johnson, J.M., Howell, A.S., Zyla, T.R., and Lew, D.J. (2008). Symmetry-breaking polarization driven by a Cdc42p GEF-PAK complex. *Curr. Biol.* **18**, 1719–1726.
- Li, R., and Bowerman, B. (2010). Symmetry breaking in biology. *Cold Spring Harb. Perspect. Biol.* **2**, a003475.
- Martin-Belmonte, F., and Perez-Moreno, M. (2011). Epithelial cell polarity, stem cells and cancer. *Nat. Rev. Cancer* **12**, 23–38.
- McCaffrey, L.M., and Macara, I.G. (2009). Widely conserved signaling pathways in the establishment of cell polarity. *Cold Spring Harb. Perspect. Biol.* **1**, a001370.
- Meitinger, F., Boehm, M.E., Hofmann, A., Hub, B., Zentgraf, H., Lehmann, W.D., and Pereira, G. (2011). Phosphorylation-dependent regulation of the F-BAR protein Hof1 during cytokinesis. *Genes Dev.* **25**, 875–888.
- Meitinger, F., Richter, H., Heisel, S., Hub, B., Seufert, W., and Pereira, G. (2013). A safeguard mechanism regulates Rho GTPases to coordinate cytokinesis with the establishment of cell polarity. *PLoS Biol.* **11**, e1001495.
- Mortensen, E.M., McDonald, H., Yates, J., 3rd, and Kellogg, D.R. (2002). Cell cycle-dependent assembly of a Gin4-septin complex. *Mol. Biol. Cell* **13**, 2091–2105.
- Nelson, W.J. (2009). Remodeling epithelial cell organization: transitions between front-rear and apical-basal polarity. *Cold Spring Harb. Perspect. Biol.* **1**, a000513.
- Noatynska, A., Tavernier, N., Gotta, M., and Pintard, L. (2013). Coordinating cell polarity and cell cycle progression: what can we learn from flies and worms? *Open Biol* **3**, 130083.
- Park, H.O., and Bi, E. (2007). Central roles of small GTPases in the development of cell polarity in yeast and beyond. *Microbiol. Mol. Biol. Rev.* **71**, 48–96.
- Park, H.O., Bi, E., Pringle, J.R., and Herskowitz, I. (1997). Two active states of the Ras-related Bud1/Rsr1 protein bind to different effectors to determine yeast cell polarity. *Proc. Natl. Acad. Sci. USA* **94**, 4463–4468.
- Pohl, C., and Jentsch, S. (2009). Midbody ring disposal by autophagy is a post-abscission event of cytokinesis. *Nat. Cell Biol.* **11**, 65–70.
- Pollarolo, G., Schulz, J.G., Munck, S., and Dotti, C.G. (2011). Cytokinesis remnants define first neuronal asymmetry in vivo. *Nat. Neurosci.* **14**, 1525–1533.
- Rothbauer, U., Zolghadr, K., Muyldermans, S., Schepers, A., Cardoso, M.C., and Leonhardt, H. (2008). A versatile nanotrap for biochemical and functional studies with fluorescent fusion proteins. *Mol. Cell. Proteomics* **7**, 282–289.
- Sherman, F. (1991). Getting started with yeast. *Methods Enzymol.* **194**, 3–21.
- Shimada, Y., Wiget, P., Gulli, M.P., Bi, E., and Peter, M. (2004). The nucleotide exchange factor Cdc24p may be regulated by auto-inhibition. *EMBO J.* **23**, 1051–1062.

- Smith, S.E., Rubinstein, B., Mendes Pinto, I., Slaughter, B.D., Unruh, J.R., and Li, R. (2013). Independence of symmetry breaking on Bem1-mediated autocatalytic activation of Cdc42. *J. Cell Biol.* 202, 1091–1106.
- Tong, Z., Gao, X.D., Howell, A.S., Bose, I., Lew, D.J., and Bi, E. (2007). Adjacent positioning of cellular structures enabled by a Cdc42 GTPase-activating protein-mediated zone of inhibition. *J. Cell Biol.* 179, 1375–1384.
- Waddle, J.A., Cooper, J.A., and Waterston, R.H. (1994). Transient localized accumulation of actin in *Caenorhabditis elegans* blastomeres with oriented asymmetric divisions. *Development* 120, 2317–2328.
- Wilcock, A.C., Swedlow, J.R., and Storey, K.G. (2007). Mitotic spindle orientation distinguishes stem cell and terminal modes of neuron production in the early spinal cord. *Development* 134, 1943–1954.
- Wloka, C., and Bi, E. (2012). Mechanisms of cytokinesis in budding yeast. *Cytoskeleton* 69, 710–726.

Proinflammatory Signaling Regulates Hematopoietic Stem Cell Emergence

Raquel Espín-Palazón,^{1,2} David L. Stachura,^{1,4} Clyde A. Campbell,¹ Diana García-Moreno,² Natasha Del Cid,¹ Albert D. Kim,¹ Sergio Candel,² José Meseguer,² Victoriano Mulero,^{2,*} and David Traver^{1,3,*}

¹Department of Cellular and Molecular Medicine, University of California, San Diego, 9500 Gilman Drive, Natural Sciences Building 6107, La Jolla, CA 92093, USA

²Departamento de Biología Celular e Histología, Facultad de Biología, Universidad de Murcia, IMIB-Arrixaca, Campus Universitario de Espinardo, Murcia 30100, Spain

³Section of Cell and Developmental Biology, University of California, San Diego, La Jolla, CA 92093, USA

⁴Present address: Department of Biological Sciences, California State University, 400 West First Street, Chico 95929-0515 CA, USA

*Correspondence: vmulero@um.es (V.M.), dtraver@ucsd.edu (D.T.)

<http://dx.doi.org/10.1016/j.cell.2014.10.031>

SUMMARY

Hematopoietic stem cells (HSCs) underlie the production of blood and immune cells for the lifetime of an organism. In vertebrate embryos, HSCs arise from the unique transdifferentiation of hemogenic endothelium comprising the floor of the dorsal aorta during a brief developmental window. To date, this process has not been replicated *in vitro* from pluripotent precursors, partly because the full complement of required signaling inputs remains to be determined. Here, we show that TNFR2 via TNF α activates the Notch and NF- κ B signaling pathways to establish HSC fate, indicating a requirement for inflammatory signaling in HSC generation. We determine that primitive neutrophils are the major source of TNF α , assigning a role for transient innate immune cells in establishing the HSC program. These results demonstrate that proinflammatory signaling, in the absence of infection, is utilized by the developing embryo to generate the lineal precursors of the adult hematopoietic system.

INTRODUCTION

In all vertebrate animals studied, the homeostasis of adult blood and immune cells is ultimately maintained by rare subsets of hematopoietic stem cells (HSCs) (Kondo et al., 2003). During a brief window during embryonic development, these HSCs arise *de novo* from hemogenic endothelium comprising the floor of the dorsal aorta (DA) (Bertrand et al., 2010a; Boisset et al., 2010; de Bruijn et al., 2000; Kissa and Herbomel, 2010) in a process that appears to be conserved among all vertebrates (Clements and Traver, 2013; Godin and Cumano, 2002). A more complete understanding of the signaling pathways that instruct HSC emergence could in principle inform *in vitro* approaches utilizing pluripotent precursors to create patient-specific HSCs (Kyba and Daley, 2003). Despite decades of efforts, this goal has not yet been achieved, in part due to an incomplete understanding of the native molecular cues needed to establish HSC fate.

One known requirement for HSC emergence is signaling through the Notch pathway (Bigas et al., 2013). Notch regulates many forms of intercellular communication, underlying many cell-fate decisions, including key roles in embryonic patterning (Kopan and Ilagan, 2009). Although the role of Notch in the maintenance and function of adult HSCs appears to be dispensable (Bigas and Espinosa, 2012), Notch signaling is absolutely required in the embryonic specification of HSCs in both the mouse (Bigas and Espinosa, 2012) and zebrafish (Bertrand et al., 2010b). In mice, the Notch receptor Notch1 (Kumano et al., 2003) and the Notch ligand Jagged1 (Jag1) are required for HSC specification (Bigas et al., 2010). It is important to note that, because Notch signaling is also indispensable for arterial specification (Quillien et al., 2014) and because HSCs derive directly from the aortic floor, it has been difficult to distinguish whether Notch signaling regulates HSC emergence independently from its role in upstream arterial specification. Recent studies in Jag1-deficient mice have demonstrated HSC defects in the presence of normal arterial development, suggesting that these Notch requirements may be distinct and separable. Recent studies have also demonstrated that Notch signaling is required intrinsically within HSCs or their precursors (Robert-Moreno et al., 2008) via function of the Notch1 receptor (Hadland et al., 2004), suggesting that Jag1 may be a specific ligand of Notch1 in the specification of HSCs.

Tumor necrosis factor α (TNF α) is a powerful proinflammatory cytokine that plays a pivotal role in the regulation of inflammation and immunity. TNF α exerts its functions via engagement of one of two specific cell surface receptors (TNFRs), namely the 55 kDa TNFR1 (also known as TNFRSF1A) and the 75 kDa TNFR2 (also known as TNFRSF1B) (Shalaby et al., 1990). TNFR1 is expressed in most cell types, whereas TNFR2 is restricted to immune and endothelial cells (Aggarwal, 2003). Whereas TNF α signaling regulates aspects of adult hematopoiesis (Mizrahi and Askenasy, 2014), a potential role in the developmental specification of HSCs has not been addressed. However, it has been reported that TNF α and its receptors are highly expressed in the murine yolk sac and fetal liver, suggesting a possible role for this inflammatory cytokine in embryonic hematopoiesis (Kohchi et al., 1994).

Nuclear factor-kappa B (NF- κ B) is a ubiquitous, inducible transcription factor that is activated by a diverse number of stimuli, including TNF α (Ahn and Aggarwal, 2005; Brown et al., 2008). A

multitude of downstream targets, as well as upstream inducers, position NF- κ B as a general sensor of cell stress. TNF α signaling through TNFR2 is a well-known activator of NF- κ B (Aggarwal et al., 2012; Faustman and Davis, 2010). TNF α activates NF- κ B through its canonical pathway, in which I κ Bs (NF- κ B inhibitors) are phosphorylated, ubiquitinated, and degraded, releasing NF- κ B dimers that then translocate to the nucleus to bind specific NF- κ B DNA binding sites to activate gene expression (Brown et al., 2008). A direct role of NF- κ B in HSCs has not been extensively studied, although recent reports indicate that NF- κ B positively regulates the transcription of genes involved in the maintenance and homeostasis of hematopoietic stem and progenitor cells (HSPCs) (Stein and Baldwin, 2013), as well as their microenvironmental interactions (Zhao et al., 2012). Whether or not NF- κ B is important in HSC emergence has not been investigated.

TNF α and TNFRs (Tnfa and Tnfrs utilizing zebrafish nomenclature) are well conserved in all vertebrate organisms (Wiens and Glenney, 2011), and we previously demonstrated that zebrafish Tnfa interacts with Tnfr1 and Tnfr2 (Espín et al., 2013). Recent studies in the zebrafish indicate that zebrafish Tnfa functions as a proinflammatory cytokine by activating endothelial cells (Roca et al., 2008). Additionally, the genetic inhibition of Tnfrs identified an essential role for Tnfa signaling in the development and maintenance of endothelial cells (Espín et al., 2013). Because HSCs arise from hemogenic endothelial cells, we queried whether TNF signaling plays a role in HSC emergence. In the present study, we demonstrate a previously unappreciated requirement for TNF signaling in the generation of HSCs. We also show that NF- κ B is active in nascent HSCs and that this activation is essential for HSC emergence. Finally, we identify primitive neutrophils as a key source of Tnfa, assigning these cells a previously unidentified role in HSC development. In summary, we report an important role for inflammatory signaling in the birth of the adult hematopoietic system that is mediated by the proinflammatory cytokine Tnfa, the inflammatory transcription factor NF- κ B, and the Notch signaling pathway under nonpathogenic conditions.

RESULTS

Tnfa Signaling through Tnfr2 Is Required for Definitive, but Not Primitive, Hematopoiesis

We previously demonstrated that Tnfa is required for embryonic blood vessel development (Espín et al., 2013). Because HSCs are generated from arterial vessels in the embryo (Bigas et al., 2013), we investigated whether this proinflammatory cytokine also played a role in HSC development. To address this question, we isolated *kdr*⁺ endothelial cells by fluorescence-activated cell sorting (FACS) from 26 hr postfertilization (hpf) transgenic *kdr*:*mCherry* embryos and performed quantitative PCR (qPCR) for *tnfr1* and *tnfr2*. Both transcripts were enriched in these cells compared to the whole embryo (Figure S1A available online). Sorted cells expressed high levels of endogenous *kdr* and were negative for the muscle-specific *myod* gene, demonstrating the purity of the sorted cells (Figure S1B). To investigate whether Tnfa signaling was required for HSC specification, we performed loss-of-function experiments for Tnfa and its two receptors, Tnfr1 and Tnfr2, utilizing specific antisense morpholinos (MOs) (Espín et al., 2013). In the zebrafish embryo, HSCs can be visualized

along the axial vessels by expression of *cmyb* using whole-mount in situ hybridization (WISH) (Burns et al., 2005). The number of *cmyb*⁺ cells in or near the floor of the DA was significantly reduced in Tnfa- and Tnfr2-deficient embryos compared with their wild-type (WT) siblings (Figures 1A and 1B). However, loss of Tnfr1 showed no effect on HSC number, and its simultaneous depletion with Tnfr2 was not significantly different than loss of Tnfr2 alone (Figures 1A and 1B), indicating that the action of Tnfa through Tnfr2, but not Tnfr1, is important in HSC development. This result was supported by quantitation of *cd41:eGFP*⁺ HSPCs (Bertrand et al., 2008) using flow cytometry, which were significantly decreased in Tnfr2- and Tnfa- deficient fish at 3 days postfertilization (dpf) (Figure 1C).

To further confirm the reduction of HSCs in Tnfr2- and Tnfa-deficient embryos, we directly visualized emerging HSCs from the floor of the DA in *kdr*:*mCherry*; *cmyb*:*GFP* double transgenic embryos (Bertrand et al., 2010a) at 48 hpf by confocal microscopy (Figure 1D). Consistent with the results above, the number of double-positive *kdr*⁺; *cmyb*⁺ HSCs in the floor of the DA was reduced ~50% when compared to control embryos (Figures 1D and 1E), unaffected in Tnfr1 deficient embryos, and showed a similar 50% decrease in Tnfr1+Tnfr2 double-depleted embryos (Figure S1C). These reductions could be due to a defect in the initial specification of HSCs or in their subsequent maintenance. To distinguish between these possibilities, we performed WISH for the nascent HSC marker *runx1* at earlier time points. Both Tnfr2- and Tnfa- deficient embryos showed significant reduction in the number of *runx1*⁺ cells in the aortic floor at 24, 28, and 36 hpf (Figures S1D and S1E), indicating that the functions of Tnfa and Tnfr2 are important during the earliest steps of HSC specification.

We next examined subsequent developmental stages for possible roles of Tnfa in the maintenance of nascent HSCs. To determine whether Tnf receptor expression is modulated following HSC specification, we purified *kdr*⁺; *cmyb*⁺ endothelial cells and *kdr*⁺; *cmyb*⁺ HSCs from 48 hpf *kdr*:*mCherry*; *cmyb*:*GFP* embryos by FACS. qPCR analysis showed that, whereas *tnfr1* mRNA levels were similar in HSCs and endothelial cells, *tnfr2* transcripts markedly increased in HSCs (Figure 1F). As this result suggested that Tnfr2 may play a role in HSC maintenance, we analyzed changes in HSC number in individual embryos over time. The number of *cmyb*⁺; *kdr*⁺ cells in WT animals expanded between 36 and 48 hpf, whereas Tnfr2- or Tnfa-deficient siblings showed similar numbers of HSCs at either time point (Figures 1G and 1H). Together, these results suggest that Tnfa signaling through Tnfr2 is important both in the first steps of HSC specification and in their subsequent maintenance following emergence from the aortic endothelium. Finally, we examined later larval stages by monitoring the expression of *rag1* and *lck*, two genes expressed in developing thymocytes (Langenau et al., 2004) because the T cell lineage derives exclusively from HSCs (Bertrand et al., 2008; Gering and Patient, 2005). Expression of *rag1* was completely or nearly absent, respectively, in Tnfr2- and Tnfa-deficient animals at 4 dpf (Figure 1I). However, the thymic anlage developed normally in all morphants, assessed by the expression of the thymic epithelial marker *foxn1* (Figure 1J). These results were further verified utilizing *lck*:*eGFP* transgenic animals to track T cell development (Langenau et al., 2004).

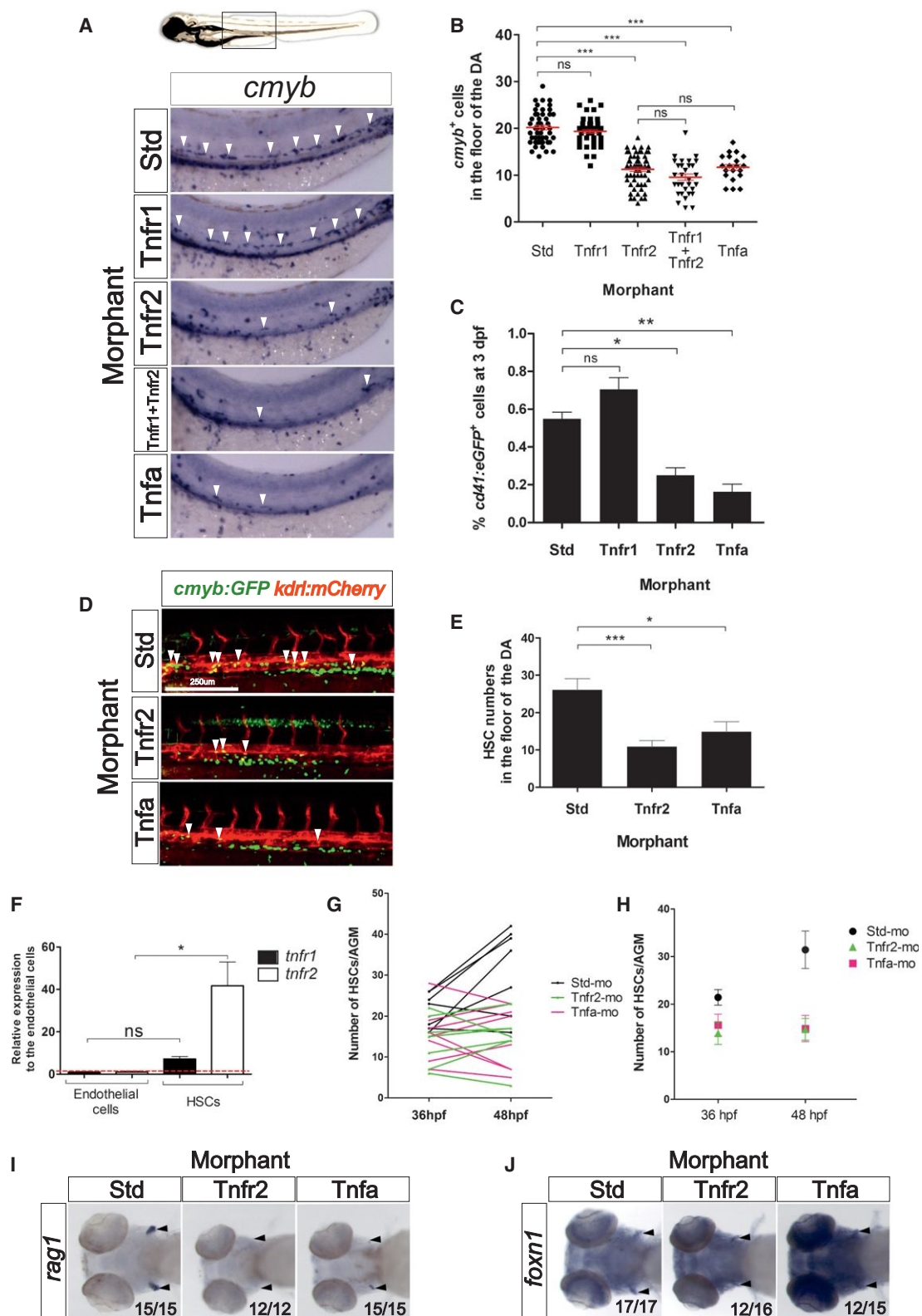


Figure 1. Tnfa and Tnfr2 Are Required for HSC Generation

(A) Standard control (Std), Tnfr1, Tnfr2, Tnfa, Tnfr1, and Tnfr2 morphants were examined by WISH for *cmyb* expression in the aortic floor at 48 hpf. White arrowheads denote *cmyb*⁺ HSPCs.

(legend continued on next page)

T cells were absent in *Tnfr2*- and *Tnfa*- deficient larvae at 4 dpf, whereas *Tnfr1*-deficient siblings showed normal T cell development (Figure S1F). Together, these results indicate that *Tnfa* signals via *Tnfr2* and that this signaling pathway is important both for early specification and subsequent maintenance of HSC fate, such that the lineage is apparently lost by 4 dpf.

To further dissect the role of *Tnfa* signaling in hematopoiesis, we assessed whether *Tnfa* and its receptors were required for the first waves of hematopoiesis, commonly referred to as “primitive” due to the transience of these cells and lack of upstream multipotent progenitors. In zebrafish, primitive hematopoiesis generates macrophages, neutrophils, and erythrocytes. The expression of *csf1ra*, a specific marker of macrophages (Herbomel et al., 2001), was unaffected in *Tnfa*-, *Tnfr1*-, and *Tnfr2*-deficient embryos at 24 hpf (Figure S1G). Additionally, primitive neutrophils were unaffected at 30 hpf, as assayed using transgenic *mpx:eGFP* animals (data not shown). Similarly, primitive erythropoiesis, assessed by expression of the erythroid-specific transcription factor *gata1a* at 24 hpf, was unaffected in morphant embryos (Figure S1G). Overall, these results indicate that *Tnfa* signaling is dispensable for primitive hematopoiesis and indispensable for definitive hematopoiesis in the zebrafish embryo.

Tnfr2- and Tnfa-Deficient Embryos Display Normal Vasculogenesis

Because HSCs originate in arterial vessels, many mutants with vascular or arterial specification defects also have hematopoietic defects (Bigas and Espinosa, 2012). No vascular abnormalities were observed in *Tnfr2*- or *Tnfa*-deficient embryos at 24 hpf when assayed by WISH for the endothelial marker *kdrl* at the MO doses used in this study (Figure 2A), and circulation was normal (*gata1:DsRed*⁺, red blood cells) but reduced numbers of HSPCs and thrombocytes (*cd41:eGFP*⁺) at 3 dpf (Figure 2B). These results suggest that the functions of *Tnfr2* and *Tnfa* are required specifically during HSC development independently of their role in developing vasculature. Thus, we could uncouple the vascular defects previously described for *Tnfr2* (Espin et al., 2013) from its effects on HSC development using lower doses of *Tnfr2* MO.

To address whether HSC defects in *Tnfr2*- and *Tnfa*-deficient animals were a consequence of impaired arterial specification, we performed WISH for the arterial markers *efnb2a*, *dlc*,

notch1b, and *notch3* (Lawson et al., 2001) in morphant embryos at 28 hpf. We observed no alterations in transcript levels when compared to control siblings (Figure 2C). Taken together, these data indicate that *Tnfa* signaling through *Tnfr2* is specifically required for HSC development.

Tnfr2 Is Intrinsically Required for HSC Development

Because *Tnfr2* is expressed in endothelial cells (Figure S1A), we hypothesized that *Tnfr2* is intrinsically required within the vascular lineage for HSC development. To test this hypothesis, we generated a transgenic zebrafish line in which the WT form of *tnfr2* is upregulated via induction of the Gal4 transcriptional transactivator. HSC development was observed by confocal microscopy following overexpression of *Tnfr2* specifically within the vasculature in *fli1a:Gal4; UAS:RFP; cmyb:GFP; UAS:tnfr2* animals. The number of RFP⁺GFP⁺ HSCs in quadruple transgenic embryos was significantly increased compared to their *Tnfr2*⁻ siblings (Figures 2D and 2E), demonstrating that *Tnfr2* activity induces or supports the HSC program following targeted expression to the vasculature.

To verify that the loss of HSCs in *Tnfr2* morphants was not due to the apoptosis of endothelial cells, we performed a TUNEL assay and immunohistochemistry for GFP in *kdrl:GFP* embryos injected with *Tnfr2* MO. Analysis of endothelial cells by confocal microscopy at 28 hpf indicated that loss of *Tnfr2* caused no increased apoptotic endothelial cells within the DA (Figure S2A), even though there was an increase in apoptotic nonendothelial cells. As a positive control for apoptosis in control animals, we imaged the lens of the eye (Cole and Ross, 2001) (Figure S2B). We also performed WISH for *runx1* in the same experiment to verify the reduction of HSCs in these embryos (Figures S2C and S2D). These results, together with the findings that there are no detectable apoptotic endothelial cells in the DA at 28 hpf (Kobayashi et al., 2014) indicate that the HSC specification defect in *Tnfr2*-deficient embryos is not caused by apoptosis induced by alterations of *Tnfr1*/*Tnfr2* ratios within the vasculature.

Tnfa Signaling Acts Upstream of Notch during HSC Specification

During Notch activation, Notch receptors are stimulated by ligands from neighboring cells, triggering the cleavage of the Notch intracellular domain (NICD), which enters the nucleus to

(B) Quantification of *cmyb*⁺ HSPCs from (A). Each dot represents total *cmyb*⁺ cells per embryo. The mean \pm SEM for each group of embryos is shown in red.
(C) *cd41:eGFP* transgenic embryos were injected with Std, *Tnfr1*, *Tnfr2*, and *Tnfa* MOs and subjected to flow cytometric analysis at 3 dpf. Each bar represents the percentage of *cd41:GFP*⁺ cells in each sample and is the mean \pm SEM of three to seven independent samples of five embryos each.
(D) Maximum projections of 48 hpf *cmyb:GFP; kdrl:mCherry* double-transgenic embryos injected with Std, *Tnfr2*, and *Tnfa* MOs. Arrowheads denote *cmyb*⁺, *kdrl*⁺ HSCs along the DA. All views: anterior to left.
(E) Enumeration of *cmyb*⁺, *kdrl*⁺ HSCs shown in (D). Bars represent mean \pm SEM of Std (n = 13), *Tnfr2* (n = 13), and *Tnfa* (n = 8) morphants.
(F) *cmyb*⁻, *kdrl*⁺ endothelial cells and *cmyb*⁻, *kdrl*⁺ HSCs were isolated from *cmyb:GFP; kdrl:mCherry* transgenic fish by FACS at 48 hpf and examined for expression of *tnfr1* and *tnfr2*. Bars represent mean \pm SEM of two biological replicates.
(G) Confocal tracking of HSC numbers in the floor of the DA from individual *cmyb:GFP; kdrl:mCherry* transgenic animals at 36 and 48 hpf following depletion of *Tnfr2* or *Tnfa* compared to standard control morphants.
(H) Means \pm SEM of *cmyb*⁺ cell numbers from (G).
(I and J) WISH for the T lymphocyte and thymic epithelial markers *rag1* (I) and *foxn1* (J) (black arrowheads), respectively, in *Tnfr2* and *Tnfa* morphants compared to Std controls at 4 dpf. All views are ventral, with anteriors to left. Numbers represent embryos with displayed phenotype; ns, not significant; *p < 0.05, **p < 0.01, and ***p < 0.001.
See also Figure S1.

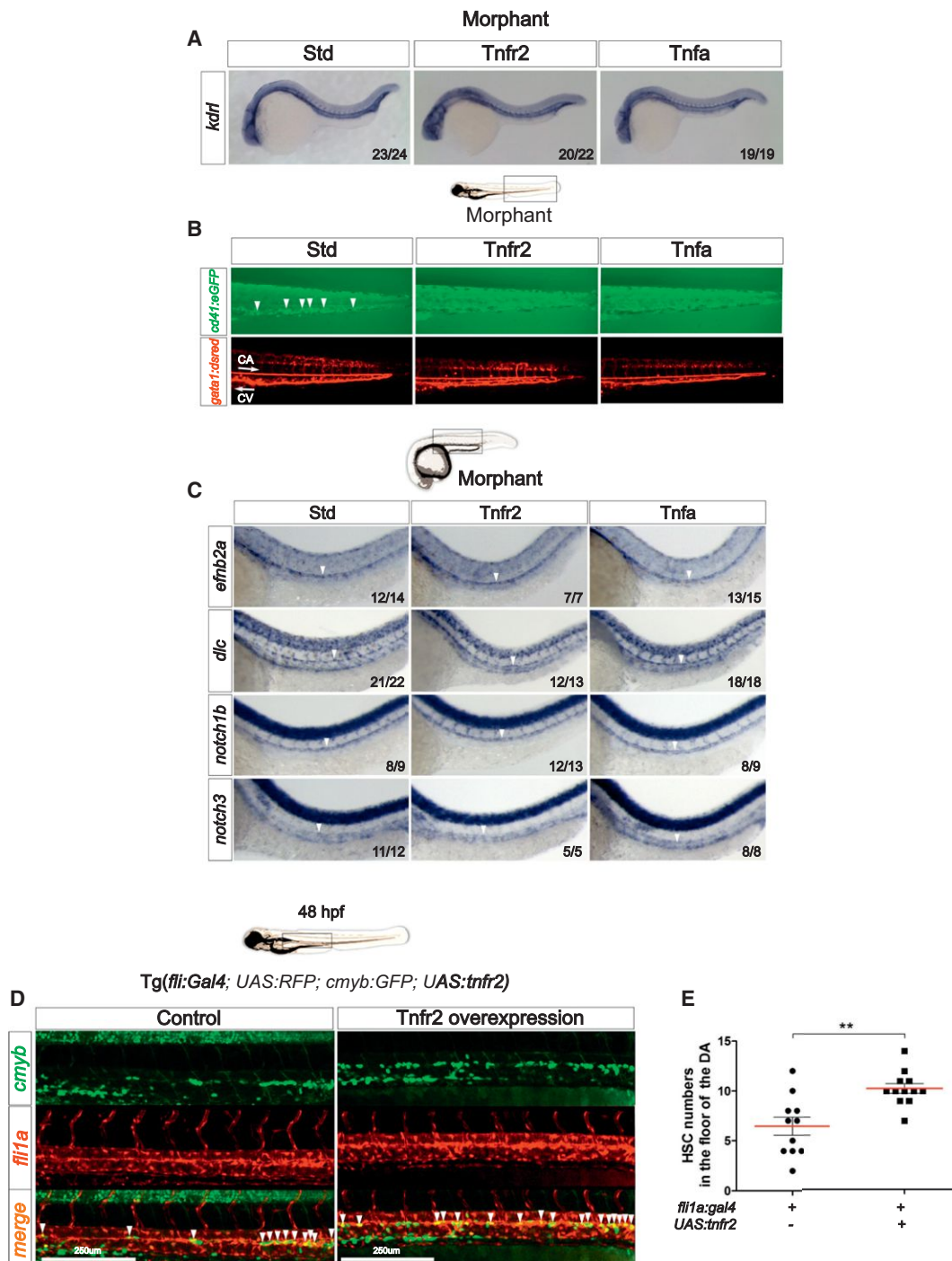


Figure 2. Signaling through Tnfr2 Regulates HSC Development Independently of Its Role in Vascular Formation

(A) Std, Tnfr2, and Tnfa morphants were interrogated by WISH for the expression of *kdr1* at 24 hpf.

(B) *cd41:eGFP*; *gata1a:dsred* double-transgenic embryos were injected with Std, Tnfr2, and Tnfa MOs and visualized at 3 dpf. Arrowheads indicate *cd41:eGFP*⁺ HSPCs in the CHT located between the caudal artery (CA) and caudal vein (CV). Arrows indicate blood flow direction.

(C) Expression of the arterial markers *efnb2a*, *dlc*, *notch1b*, and *notch3* in Std, Tnfr2, and Tnfa morphants analyzed by WISH at 28 hpf. Arrowheads denote the CA.

(D) Maximum projections of *fil1a:Gal4*; *UAS:tnfr2*; *cmyb:eGFP*; *kdr1:mCherry* transgenic embryos at 48 hpf. Region shown includes the DA, and arrowheads denote *cmyb*⁺; *kdr1*⁺ HSCs.

(E) Enumeration of *cmyb*⁺; *kdr1*⁺ HSCs shown in (D). Each dot is the number of *kdr1*⁺; *cmyb*⁺ cells per embryo. Means \pm SEM for each group is shown in red.

**p < 0.01. All views are lateral, with anteriors to the left. Numbers in panels represent larvae with indicated phenotype.

See also Figure S2.

function as transcription factor essential for cell fate decisions (Lai, 2004). There are four Notch receptors (Notch1a, 1b, 2, and 3), five Delta family ligands (Dla, Dlb, Dlc, Dld, and Dll4) and three Jagged ligands (Jagged 1a, Jagged 1b, and Jagged 2) in zebrafish. Because TNF α activates the Notch pathway in certain contexts (Fernandez et al., 2008; Wang et al., 2013), we queried whether signaling through Tnfr2 may similarly activate Notch signaling to specify HSCs. We performed loss-of-function experiments for Tnfr2 and Tnfa in transgenic *tp1:eGFP* animals, in which GFP is expressed by cells having recently experienced Notch signaling (Parsons et al., 2009). Consistent with our other findings, the depletion of either Tnfa or Tnfr2 led to a 2-fold reduction in *tp1:eGFP*⁺; *kdr1:mCherry*⁺ HSPCs in the aortic floor at 26 hpf (Figure 3A, arrowheads, and Figure 3B). These observations indicate that Tnfr2 signaling is upstream of Notch signaling during HSC specification.

If Notch signaling is indeed required downstream of Tnfr2 function for HSC specification, then ectopic expression of the Notch1a intracellular domain (NICD1a) should rescue the lack of HSCs in Tnfr2- and Tnfa-deficient embryos. We performed two different experiments to address the timing and tissue specificity of this Tnfa-dependent Notch requirement. To provide temporal control of NICD1a induction, we utilized inducible *hsp70:Gal4*; *UAS:NICD1a-myc* double-transgenic embryos, which express NICD1a under the control of the inducible Gal4 system. Induction of NICD1a at 18 hpf rescued the depletion of *runx1*⁺ HSCs at 28 hpf along the DA in both Tnfa and Tnfr2 morphants (Figure 3C). We then enforced the expression of NICD1a within endothelial cells utilizing *kdr1:Gal4*; *UAS:NICD1a-myc* double-transgenic embryos that had been injected with Tnfr2 or Tnfa MOs. Endothelial expression of NICD1a restored *runx1*⁺ cells along the aortic floor (Figure 3D), indicating that TNF signaling activates the Notch pathway within hemogenic endothelium to specify HSC fate.

Tnfa Induces Jag1a within Endothelial Cells to Promote HSC Specification through Notch1a

We next investigated potential mechanisms by which Tnfa and Tnfr2 induced Notch activation. Due to the fact that Tnfa signaling has been reported to induce or inhibit the expression of specific Notch ligands (Fernandez et al., 2008; Sainson et al., 2008), we analyzed expression of the eight zebrafish Notch ligands within purified *kdr1*⁺ endothelial cells from Tnfr2-deficient embryos. Only *jag1a* expression was downregulated in Tnfr2 morphants relative to controls (Figure 4A). Using a *fli1a:Gal4* driver to enforce expression of Tnfr2 specifically within the vasculature, we examined Notch ligand expression in *fli1a:Gal4*; *UAS:tnfr2* animals by qPCR (Figure 4B). We detected a 20-fold increase of *tnfr2* in *UAS:tnfr2*⁺ compared to *UAS:tnfr2*⁻ embryos (Figure 4B). Consistent with our previous results, only *jag1a* mRNA levels were increased following the enforced expression of Tnfr2 (Figure 4B).

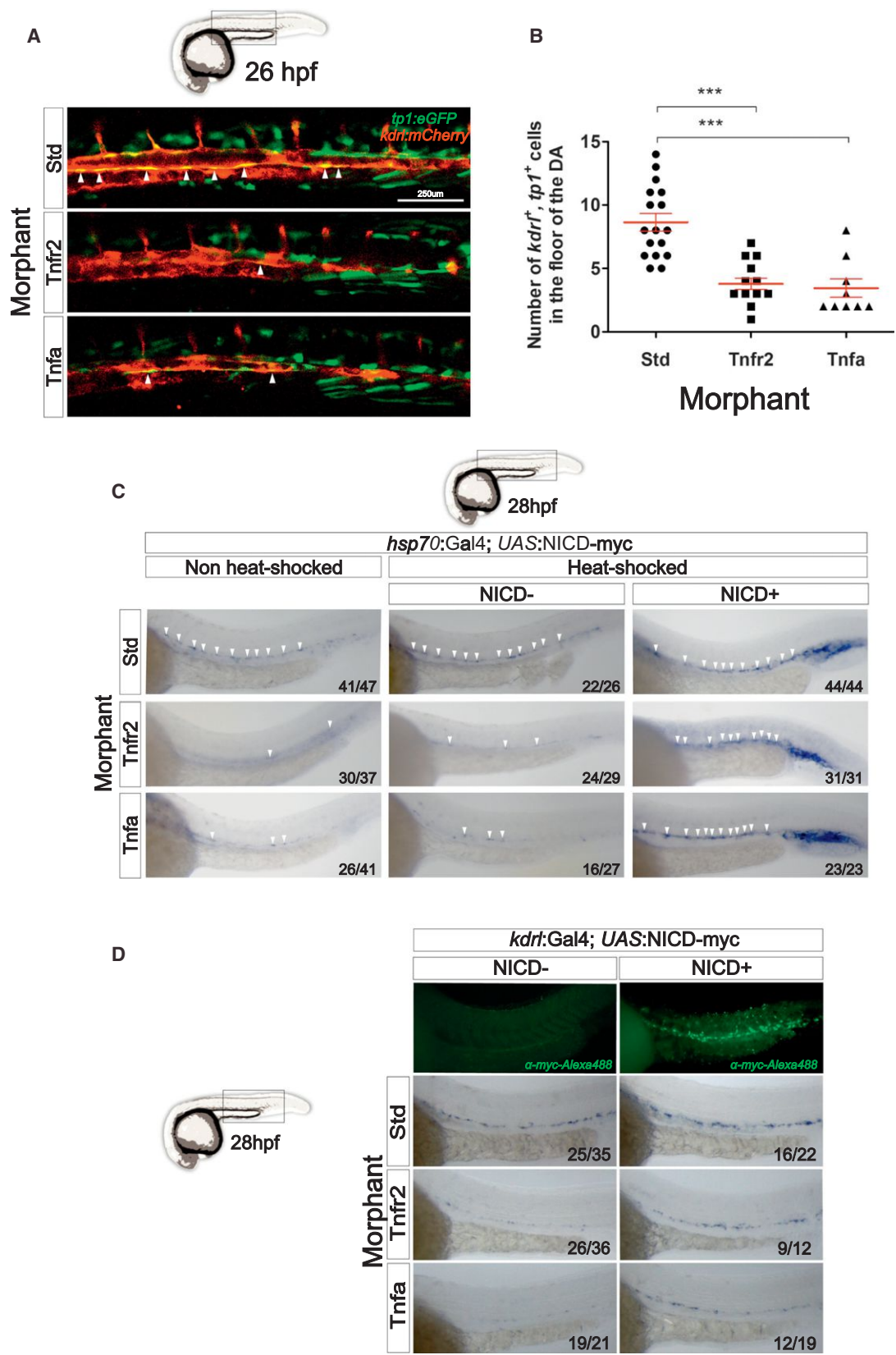
Interestingly, Jag1 is required for the generation of definitive hematopoietic cells in mice but is dispensable for arterial development. A potential role for Jag1 in zebrafish HSPC development has not been addressed. Two paralogues of the single JAG1 human gene are present in the zebrafish genome: *jag1a* and *jag1b*. Because only *jag1a* levels were modulated by Tnfr2, we per-

formed loss-of-function experiments with this gene. Loss of *jag1a* led to decreased HSC numbers as analyzed by *runx1* expression along the DA (Figure 4C). However, specification of aortic fate was normal, as *efnb2a* and *dlc* levels were unperturbed (Figure 4C). To further verify that Tnfr2 and Jag1a were in the same genetic pathway, we performed synergy studies by coinjecting low doses of Tnfr2 and Jag1a MOs simultaneously. Aortic *runx1*⁺ cells were significantly reduced in Tnfr2- and Jag1a- double-deficient embryos compared to single-deficient embryos (Figure 4D). Tnfr2 function thus lies genetically upstream of *jag1a* during HSC specification. To investigate potential Jag1a-presenting cells, we isolated *cmyb*⁻, *kdr1*⁺ endothelial cells and *cmyb*⁺, *kdr1*⁺ HSCs for qPCR analysis of *jag1a* at 48 hpf. *jag1a* transcripts were 4-fold more abundant in endothelial cells than in HSCs (Figure S3), suggesting that Notch signaling in HSCs or hemogenic endothelium is activated by neighboring Jag1a⁺ endothelial cells.

We next investigated which of the four Notch receptors were downstream of Jag1a during HSC induction. In the mouse, Notch1 is required within HSCs or their lineal precursors to instruct HSC fate. We therefore focused upon the two zebrafish orthologs of human NOTCH1, Notch1a and Notch1b. To investigate whether either receptor functioned downstream of Tnfr2 to specify HSCs, we performed synergy experiments by coinjecting low doses of Tnfr2 MO with morpholinos against Notch1a or Notch1b. Only the simultaneous depletion of Tnfr2 and Notch1a, but not Tnfr2 and Notch1b, led to a statistically significant decrease in *runx1* expression compared to single morphants at 28 hpf (Figure 4E). This finding suggests that Notch1a serves as the Notch receptor for Jag1a to specify HSC fate downstream of Tnfr2.

The Proinflammatory Transcription Factor NF- κ B Is Active in Emerging HSCs

Activation of TNF receptors by ligand binding leads to the recruitment of adaptor proteins that trigger NF- κ B activation (Aggarwal et al., 2012). Moreover, the induction of *Jag1* transcription by Tnfa in murine endothelial cells is NF- κ B dependent (Johnston et al., 2009). Interestingly, NF- κ B (as well as Tnfr2 and Jag1) is necessary for embryonic vessel development (Santoro et al., 2007). These lines of evidence suggested that NF- κ B could have a previously unappreciated role in HSC specification, prompting us to examine its role in HSC development. We utilized an NF- κ B activation reporter transgenic line (Kanter et al., 2011) in combination with the *kdr1:mCherry* transgene to perform confocal analysis of the DA at different time points. Interestingly, we observed NF- κ B^{high} cells in the floor of the DA at 24 hpf, typically in pairs and in direct contact with each other (Figure 5A). We also observed NF- κ B⁺ cells along the roof of the DA but at a much lower frequency than in the floor (data not shown). NF- κ B⁺, *kdr1*⁺ cells remained visible at 30 hpf (Figure 5A) and underwent endothelial-to-hematopoietic transition (EHT) (Movie S1), a characteristic feature of emerging HSCs. To further evaluate whether HSCs had increased NF- κ B activation compared to their surrounding endothelial neighbors, *kdr1*⁺; *cmyb*⁺ HSCs and *kdr1*⁺; *cmyb*⁻ endothelial cells were isolated from 48 hpf *kdr1:mCherry*; *cmyb:GFP* embryos by FACS for qPCR analyses. Whereas endothelial cells had 20- to 30-fold induction of the NF- κ B response genes interleukin 1 beta



(legend on next page)

(*il1b*) and nuclear factor of kappa light polypeptide gene enhancer in B cells inhibitor alpha a (*ikbaa*) relative to whole-embryo expression, HSCs displayed 300- and 2,300-fold increases in *il1b* and *ikbaa*, respectively (Figure 5B). Immunohistochemistry for the NF- κ B subunit p65 in *kdr1:mCherry* embryos showed that, although p65 was detected in the cytoplasm of every cell as expected, it was more intense in the pronephros (Figures 5C and 5D, yellow asterisks), in the DA, and in cells potentially undergoing the endothelial to hematopoietic transition in the aortic floor (Figures 5C and 5D, arrow). These results indicate that NF- κ B activation is a characteristic feature of emerging HSCs.

Multiple lines of evidence support the integration of the Notch and NF- κ B signaling pathways during the differentiation of various cell types (Ang and Tergaonkar, 2007; Cao et al., 2011; Espinosa et al., 2010; Espinosa et al., 2003; Shin et al., 2006; Song et al., 2008). For this reason, we investigated whether NF- κ B⁺ cells in the floor of the DA also had active Notch signaling, utilizing double-transgenic *tp1:nlsCherry; NFKB:GFP* animals to simultaneously visualize respective Notch and NF- κ B activation. NF- κ B⁺ cells in the floor of the DA were also *tp1*⁺ (Figure 5E). No NF- κ B⁺, *tp1*⁻ cells were found in the floor of the DA, suggesting that Notch is (or was previously) active in NF- κ B⁺ HSPCs.

NF- κ B Activation Is Required for HSC Specification and Acts Downstream of Tnfr2

To determine whether NF- κ B function is required for HSC emergence, we developed a Tg(*UAS:dn-ikbaa*) transgenic animal that functions as a dominant-negative inhibitor of NF- κ B (Figures S4A and S4B). Similar truncation constructs have been utilized in vitro to inhibit NF- κ B activation (Abbas and Abu-Amer, 2003). At 6 hr post-heat-shock in *hsp70:Gal4; UAS:dn-ikbaa* animals, *dn-ikbaa* mRNA levels were detected in *dn-ikbaa*⁺, but not in *dn-ikbaa*⁻, siblings (Figure S4C). qPCR for the NF- κ B response gene *il1b* in FACS-purified *fli1a*⁺ endothelial cells showed significant downregulation in the *dn-ikbaa*⁺ embryos compared to their *dn-ikbaa*⁻ siblings (Figures S4D and S4E). Lipopolysaccharide (LPS) challenge of WT embryos produced a significant increase in *il1b* expression compared to PBS-injected controls, as previously described (van der Vaart et al., 2013), but not in *dn-ikbaa*⁺ embryos (Figures S4F and S4G), indicating that *dn-ikbaa*⁺ embryos are unable to trigger an inflammatory response through NF- κ B. These results thus demonstrate that *UAS:dn-ikbaa* embryos have impaired NF- κ B activation.

Blockade of NF- κ B function at 20 hpf in *hsp70:Gal4; UAS:dn-ikbaa* animals led to loss of HSCs at 48 hpf (Figure 6A). Loss of NF- κ B specifically within the vasculature using *fli1a:Gal4; UAS:dn-ikbaa* double-transgenic embryos also led to a depletion

of *cmyb*⁺ cells (Figure 6B). qPCR for *il1b* in FACS-purified endothelial cells showed a 3-fold decrease in Tnfr2 morphants (Figure 6C), demonstrating that NF- κ B acts downstream of Tnfr2 during HSC specification. Together, these results suggest that NF- κ B activation in hemogenic endothelium is a key event in the specification of HSCs.

Primitive Neutrophils Are the Key Source of Tnfa

In adult organisms, immune cells are the main source of TNF α , including T and B lymphocytes, macrophages, and neutrophils (Aggarwal, 2003). From 22 to 72 hpf, the temporal window over which zebrafish HSCs emerge from aortic endothelium, the only leukocytes present are primitive myeloid cells, namely macrophages and neutrophils (Herbomel et al., 1999; Le Guyader et al., 2008). Interestingly, *tnfa* expression in the zebrafish embryo was not detectable during the first 9 hr of development but was expressed before 24 hpf (Espín et al., 2013) when HSCs are initially specified (Clements and Traver, 2013). We therefore hypothesized that primitive myeloid cells were the source of Tnfa. We isolated *mpeg:GFP*⁺ primitive macrophages and *mpx:GFP*⁺ primitive neutrophils by FACS at two different time points and performed qPCR for *tnfa* (Figure 7A). Although both populations expressed *tnfa*, the highest expression was observed within the neutrophil fraction (Figure 7A). We then utilized a pu1 MO to specifically ablate both primitive myeloid lineages in vivo (Rhodes et al., 2005). pu1 MO efficacy was validated by WISH using the panleukocyte marker *i-plastin* and the neutrophil marker *mpx* at 48 hpf (Figure S5A). Following ablation of primitive myeloid cells in pu1 morphants, HSCs were enumerated by confocal microscopy of *kdr1*⁺*cmyb*⁺ cells. A 2-fold decrease in HSC number was detected in pu1 morphants compared to their control siblings (Figures 7B and 7C). To elucidate which primitive myeloid population was responsible for the decrease in HSC number, we utilized an irf8 MO (Li et al., 2011), which skews myeloid development to almost entirely neutrophilic. Loss of the macrophage lineage was confirmed in irf8 morphants by qPCR for the macrophage-specific marker *mpeg1* (Figure S5B). Surprisingly, the number of *kdr1*⁺; *cmyb*⁺ HSCs increased following loss of the macrophage lineage (Figures 7D and 7E). In agreement with our *tnfa* expression data, this result suggests that neutrophils are the key source of the Tnfa needed for HSC emergence. To test this hypothesis, we quantified *tnfa* expression levels in pu1- and irf8-deficient animals. Expression of *tnfa* was consistently decreased following loss of pu1 function and increased following loss of irf8 function (Figure 7F). In addition, although *runx1* was upregulated in irf8-deficient embryos, the simultaneous depletion of Tnfa and Irf8 led to a marked reduction in *runx1* expression, despite the elevated numbers of neutrophils present (Figure 7G). These

Figure 3. Tnfa and Tnfr2 Act Upstream of Notch Signaling during HSC Specification

(A) *tp1:eGFP; kdr1:mCherry* embryos injected with Std, Tnfr2, and Tnfa MOs were visualized at 26 hpf. Arrowheads indicate cells in the floor of the DA with active Notch signaling.
 (B) Enumeration of *tp1*⁺, *kdr1*⁺ HSCs from (A). Each dot represents the number of HSCs per embryo, and red lines indicate means \pm SEM. ***p < 0.001.
 (C) *hsp70:Gal4; UAS:NICD-myc* embryos injected with Std, Tnfr2, and Tnfa MOs were heat shocked at 18 hpf and WISH for *runx1* was performed at 28 hpf. Arrowheads denote HSCs along the DA.
 (D) *kdr1:Gal4; UAS:NICD-myc* embryos injected with Std, Tnfr2, and Tnfa MOs were analyzed by WISH for *runx1* at 28 hpf. NICD⁺ larvae were identified using anti-myc-Alexa488 antibody (top). Numbers in panels represent the numbers of larvae with indicated phenotype.

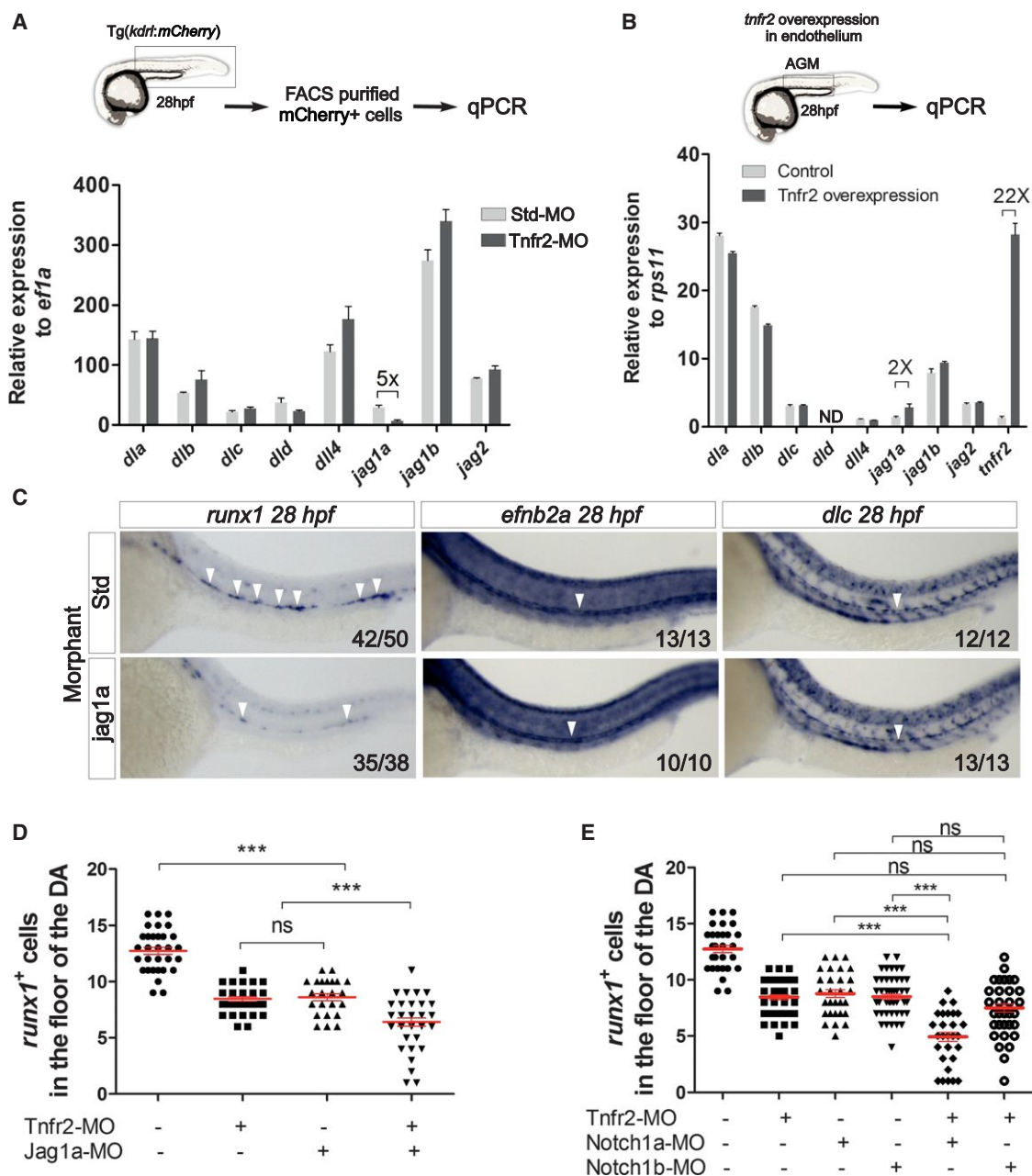


Figure 4. Tnfr2 Induces *jagged1a* in Endothelial Cells, Encouraging HSC Specification

(A) *kdr1:mCherry*⁺ cells from dissected trunks of Std or Tnfr2 morphants were purified by FACS at 28 hpf for qPCR. Levels of indicated transcripts along x axis are shown relative to the housekeeping gene *ef1a*. Bars represent means \pm SEM of duplicate samples.

(B) AGM regions from *fli1a:Gal4*; *UAS:tnfr2* embryos were dissected and subjected to qPCR for transcripts shown along x axis. Bars represent means \pm SEM of triplicate samples expression relative to the housekeeping gene *rps11*.

(C) Std (top) or Jag1a (bottom) morphants were interrogated for *runx1* expression at 26 hpf and *efnb2a* and *dlc* at 28 hpf by WISH. Numbers represent larvae with indicated phenotype.

(D) Enumeration of *runx1*⁺ cells in Tnfr2 and/or Jag1a morphants at 28 hpf.

(E) Enumeration of *runx1*⁺ cells in Tnfr2 and/or Notch1a and/or Notch1b morphants at 28 hpf. Each dot is the number of HSCs per embryo, and red lines indicate means \pm SEM (D and E) ***p < 0.001; ns, not significant.

See also Figure S3.

findings demonstrate that production of Tnfa from primitive neutrophils is critical for the specification and/or maintenance of HSC fate.

Overall, these data indicate that production of Tnfa from primitive neutrophils activates Tnfr2, upregulating the expression of Jag1a on the surface of endothelial cells. Jag1a in turn activates

Notch1a, triggering a signaling cascade whereby NF- κ B triggers a transcriptional program required for the emergence of HSCs from hemogenic endothelium (Figures S5C and S5D).

DISCUSSION

Traditionally, infection and inflammation were thought to play an indirect role in HSC homeostasis by causing increased proliferation and skewed differentiation toward microbicidal immune cell lineages (Takizawa et al., 2012). However, recent studies indicate that HSCs can respond directly to the inflammatory cytokines interferon (IFN) α/β , γ , and TNF α (Baldrige et al., 2011; King and Goodell, 2011). Additionally, there is evidence that HSCs can upregulate cytokines under stress-induced hematopoiesis (Zhao et al., 2014). Here, we examined a much earlier step in the biology of HSCs—their specification and emergence from hemogenic endothelium in the developing embryo. The emergence of HSCs from the aortic floor is transient and occurs during developmental windows when the surrounding environment is relatively sterile, whether it is in utero in mammals or within the chorion in teleosts. It is therefore surprising that a key pathway underlying the canonical response to infection and inflammation is required to generate the founders of the adult hematopoietic system. Our studies in the zebrafish demonstrate that depletion of Tnfa or its cognate receptor Tnfr2 leads to depletion of emerging HSCs. The key event elaborated by Tnfr2 appears to be activation of the Notch pathway because ectopic provision of Notch signaling rescued HSCs in the absence of Tnfa or Tnfr2 function. Although Notch signaling is required for HSC specification across vertebrate phyla, little is known regarding how this Notch event is regulated or which of the many receptors or ligands are necessary to fate HSCs from ventral aortic endothelium.

That the HSC program can be rescued in either Tnfa or Tnfr2 morphants by enforced expression of NICD1a within the vasculature demonstrates that the TNF pathway lies upstream of Notch in HSC specification. Our results suggest that signaling via Tnfr2 specifically controls Notch activation by inducing the Notch ligand *jag1a* in cells within the DA. Synergy experiments depleting Notch1a and Tnfr2 combinatorially indicate that Notch1a is likely the receptor on HSCs that binds to the Jag1a ligand presented by aortic endothelial cells. These findings are consistent with studies in the mouse embryo, where Notch1 is required cell autonomously within HSCs or their lineal precursors for their specification (Hadland et al., 2004; Kumano et al., 2003). The zebrafish Notch1a and Notch1b receptors are evolutionary paralogues of mammalian Notch1 (Kortschak et al., 2001) and are both expressed in the DA during the window of HSC emergence (Quillien et al., 2014). Our findings extend these results by demonstrating that Notch1 function is evolutionarily conserved in the specification of HSCs and provide a more detailed mechanism regarding how Notch1 may actually function in this process. Further studies will be required to determine the precise interactions between Jag1a and Notch1a and how these interactions lead to establishment of HSC fate.

In addition to its regulation of the Notch pathway, our results also suggest that Tnfa exerts its effects through NF- κ B. Although NF- κ B is known to play a key role in adult mammalian hematopoi-

esis (Gerondakis et al., 2012), a role in the embryonic emergence of HSCs has not been reported. The utilization of a *NF κ B:GFP* reporter line allowed us to image the in vivo activation of NF- κ B, indicating that this activation is required within endothelial cells of the DA for HSC emergence. Furthermore, these studies suggest that this activity is downstream of Tnfa/Tnfr2 signaling. Intriguingly, these data also demonstrate that NF- κ B⁺ cells in the floor of the DA are often positive for Notch activity when assessed along with the *tp1* Notch reporter line. Whereas recent evidence suggests that Notch1 can modulate NF- κ B activity in different cellular contexts, it remains to be determined whether one factor is epistatic to the other or if both may operate together within the hemogenic endothelium to establish HSC fate.

In this study, we have also discovered an unexpected role for neutrophils in HSC development. Whereas macrophages are involved in a broad array of developmental processes (Wynn et al., 2013), an active role for neutrophils in modulating developmental events has not been described. Here, we report for the first time that primitive neutrophils are a major source of Tnfa and that the loss of either neutrophils or Tnfa results in the loss of developing HSCs. The prevailing view that primitive myeloid cells have evolved predominantly to provide early immunity is thus likely oversimplistic. At any time point during HSC emergence, whether early during HSC specification or later during EHT, we observed ~2-fold decreases in HSC number. That the lineal descendants of HSCs, most importantly T lymphocytes, are absent by 4–5 dpf indicates that the Tnfa/Tnfr2 signaling axis is required to sustain HSC function. Collectively, our findings suggest that activation of Tnfr2 is important both in hemogenic endothelium and in maintaining nascent HSC fate. It is important to note that *tnfa* is also expressed in endothelial cells (data not shown); contribution from the endothelium may thus play a role in either or both of these processes. The means to create conditional, tissue-specific gene disruption in the zebrafish will be required to precisely address the relative importance of each source.

In conclusion, we show that TNF α , a cytokine that has become the paradigm for induction of inflammatory responses, is also key in the establishment of the hematopoietic system through its influence on HSC formation in the developing embryo. In addition to the known signaling inputs required to establish HSC fate, inflammatory signals should now be added to this list. A major challenge for the field is to integrate each of these required inputs to better understand their spatial and temporal requirements, such that this knowledge may be utilized to instruct HSC fate in vitro from human pluripotent precursors, a major unrealized goal of regenerative medicine.

EXPERIMENTAL PROCEDURES

Zebrafish Husbandry and Strains

Zebrafish embryos and adults were mated, staged, raised, and processed as described (Westerfield, 2000) and maintained in accordance with UCSD IACUC guidelines. See Extended Experimental Procedures for description of transgenic lines.

Heat-Shock Treatment

For induction of *hsp70l:Gal4*-driven transgenes, embryos were placed in E3 medium and transferred to a 38°C water bath for 45 min at noted stages.

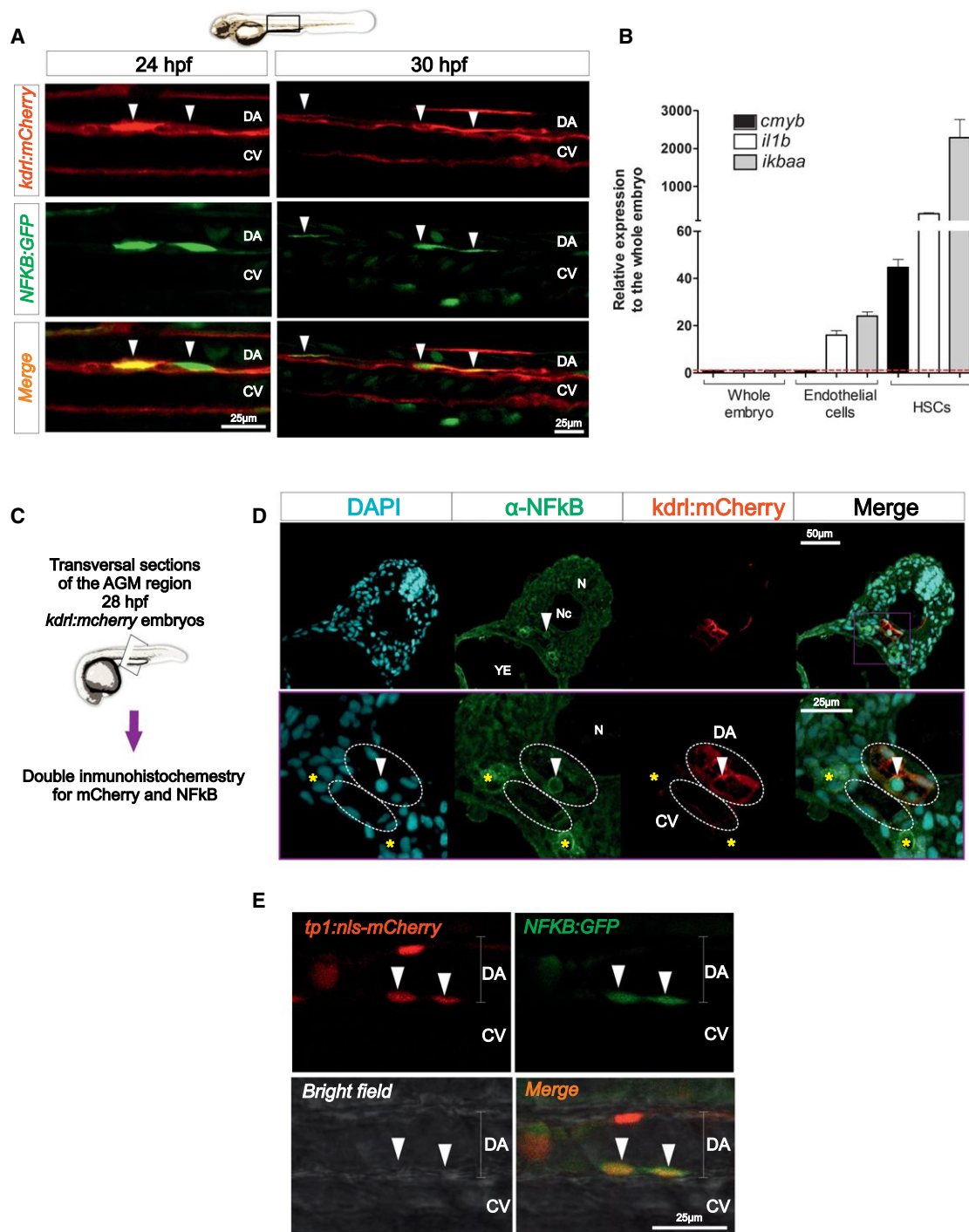


Figure 5. NF- κ B Is Active in Emerging HSCs

(A) Trunk region of *kdr1:mCherry*; *NFkB:GFP* double-transgenic animals visualized by confocal microscopy at 24 hpf (left) and 30 hpf (right). Each image is a 2 μ m z slice. Arrowheads denote HSCs.

(B) *cmyb*⁺, *kdr1*⁺ endothelial cells and *cmyb*⁺, *kdr1*⁺ HSCs were isolated by FACS at 48 hpf. Levels of the NF- κ B target genes *ikbaa* and *il1b*, as well as the HSC marker *cmyb*, are shown relative to *ef1a*. Bars represent means \pm SEM of two biological replicates.

(legend continued on next page)

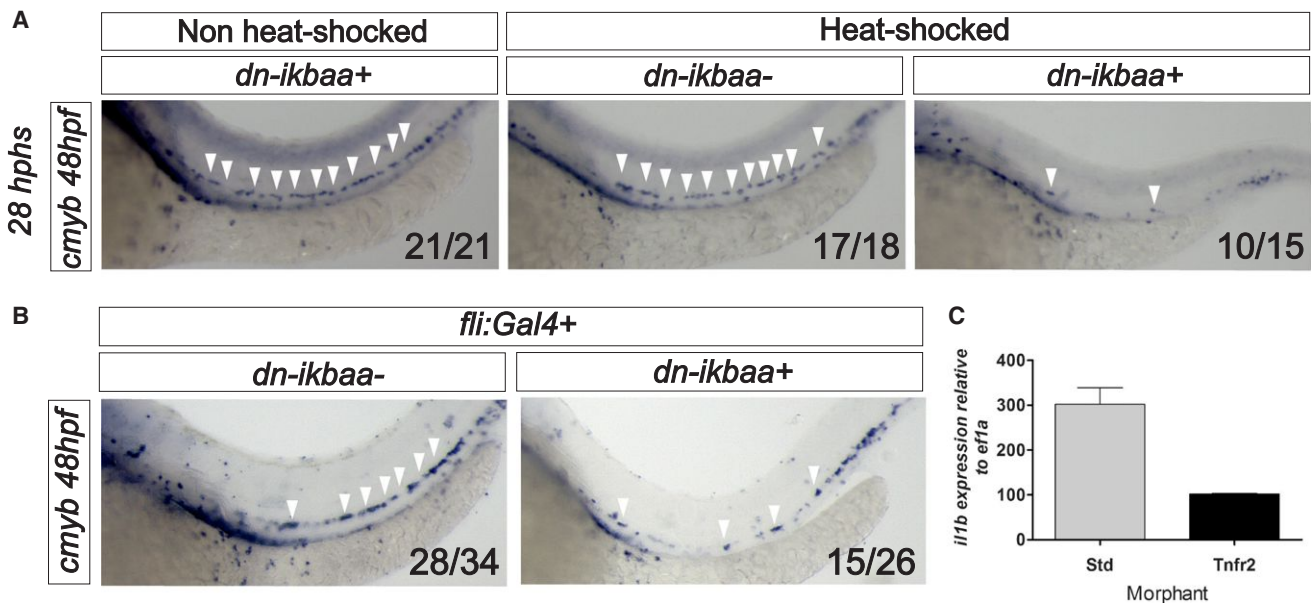


Figure 6. NF- κ B Is Required for HSC Specification and Acts Downstream of Tnfr2

(A) *hsp70:Gal4; UAS:dn-ikbaa* embryos were heat shocked at 20 hpf. WISH for *cmyb* was performed at 48 hpf.

(B) WISH for *cmyb* in *fli1a:Gal4; UAS:dn-ikbaa⁻* (left) and *fli1a:Gal4; UAS:dn-ikbaa⁺* (right) embryos. Arrowheads mark *cmyb⁺* cells along the DA.

(C) *kdr1:mCherry⁺* cells were FACS sorted from Std or Tnfr2 morphants at 28 hpf for qPCR. Levels of the NF- κ B target gene *il1b* are shown relative to *ef1a*. Bars represent means \pm SEM from duplicate samples.

See also Figure S4.

Generation of Transgenic Animals

Tg (UAS:dnnfkb1a)^{sd35} and *Tg(UAS:tnfr2)^{ums1}* embryos were generated by Tol2-mediated transgenesis via the multisite Gateway cloning system (Invitrogen). See also Extended Experimental Procedures.

Morpholino Injection

Specific antisense targeting MOs (Gene Tools) were resuspended in DEPC-treated water at 1–3 mM and injected in one-cell stage embryos. See also Extended Experimental Procedures.

Enumeration of HSCs

Animals were subjected to WISH for *runx1* and *cmyb* at noted stages, and positive cells were imaged and manually counted. Confocal microscopy was performed on *cmyb:GFP; kdr1:mCherry* double-transgenic animals (Bertrand et al., 2010a), *tp1:eGFP; kdr1:mCherry* double-transgenic animals, and NF- κ B:GFP; *kdr1:mCherry* double-transgenic animals. Z sections of the DA region were captured on a Leica SP5 microscope (Leica) using Volocity Acquisition, Visualization, and Restoration software (Improvision) and were manually counted.

Fluorescent Visualization of Blood Flow, HSPCs, and T Cells

To visualize blood flow, HSPCs, and T cells, *cd41:eGFP; gata1:dsred* embryos at 3 dpf and *lck:GFP* larvae at 4 dpf, respectively, were anesthetized in Tricaine (200 μ g/ml) and examined using a Leica MZ16FA stereomicroscope.

Flow Cytometry and FACS

Briefly, embryos were dechorionated with pronase, anesthetized in tricaine, and dissociated with liberase or triturated with a P1000 pipette. The resulting suspension was filtered with a 40 μ m cell strainer, and flow cytometric acqui-

sitions or FACS were performed on a FACS LSRII. See also Extended Experimental Procedures.

Whole-Mount RNA In Situ Hybridization

WISH was carried out as described (Thisse et al., 1993). Probes for the *gata1a*, *csfr1a*, *kdr1*, *cmyb*, *runx1*, *foxn1*, *efnb2a*, *dlc*, *notch1b*, *notch3*, and *rag1* transcripts were generated using the DIG RNA Labeling Kit (Roche Applied Science) from linearized plasmids. *dn-ikbaa* probe was generated from bp 118–933 of *dn-ikbaa* (see Figure S2). Embryos were imaged using a Leica M165C stereomicroscope equipped with a DFC295 color digital camera (Leica) and FireCam software (Leica).

Statistical Analyses

Data were analyzed by analysis of variance (ANOVA). In all figures, solid red bars denote the mean, and error bars represent SEM. * $p < 0.05$, ** $p < 0.01$, and *** $p < 0.001$; n.s., not significant; n.d., not detected.

Quantitative RT-PCR Analysis

RNA was isolated from tissues with RNeasy (QIAGEN), and cDNA was generated with qScript Supermix (Quanta BioSciences). Primers to detect zebrafish transcripts are described in Table S1. Relative expression levels of genes were calculated by the following formula: relative expression = $2^{-(Ct[\text{gene of interest}] - Ct[\text{housekeeping gene}])}$.

Immunofluorescence of NICD⁺ Animals

The immunofluorescence staining for cMyc in *hsp70:gal4; UAS:NICD-myc* zebrafish embryos was performed as previously described (Kim et al., 2014).

(C) Schematic representation of the experimental design of (D). 28 hpf *kdr1:mCherry* animals were transversally sectioned and subjected to double immunohistochemistry for mCherry (red) and NF- κ B (green). DAPI (blue) was added to visualize nuclei.

(D) Maximum projections of 1 μ m sections. Arrowhead indicates a potential HSC emerging in the DA. DA and CV are demarcated by dashed white lines. Yellow asterisks indicate pronephric ducts.

(E) *tp1:nls-mCherry; NFkB:GFP* animals were visualized by confocal microscopy at 24 hpf. Each image is a 2 μ m z slice. Arrowheads indicate HSCs. CV, caudal vein; DA, dorsal aorta; N, neural tube; Nc, notochord; YE, yolk extension.

See also Movie S1.

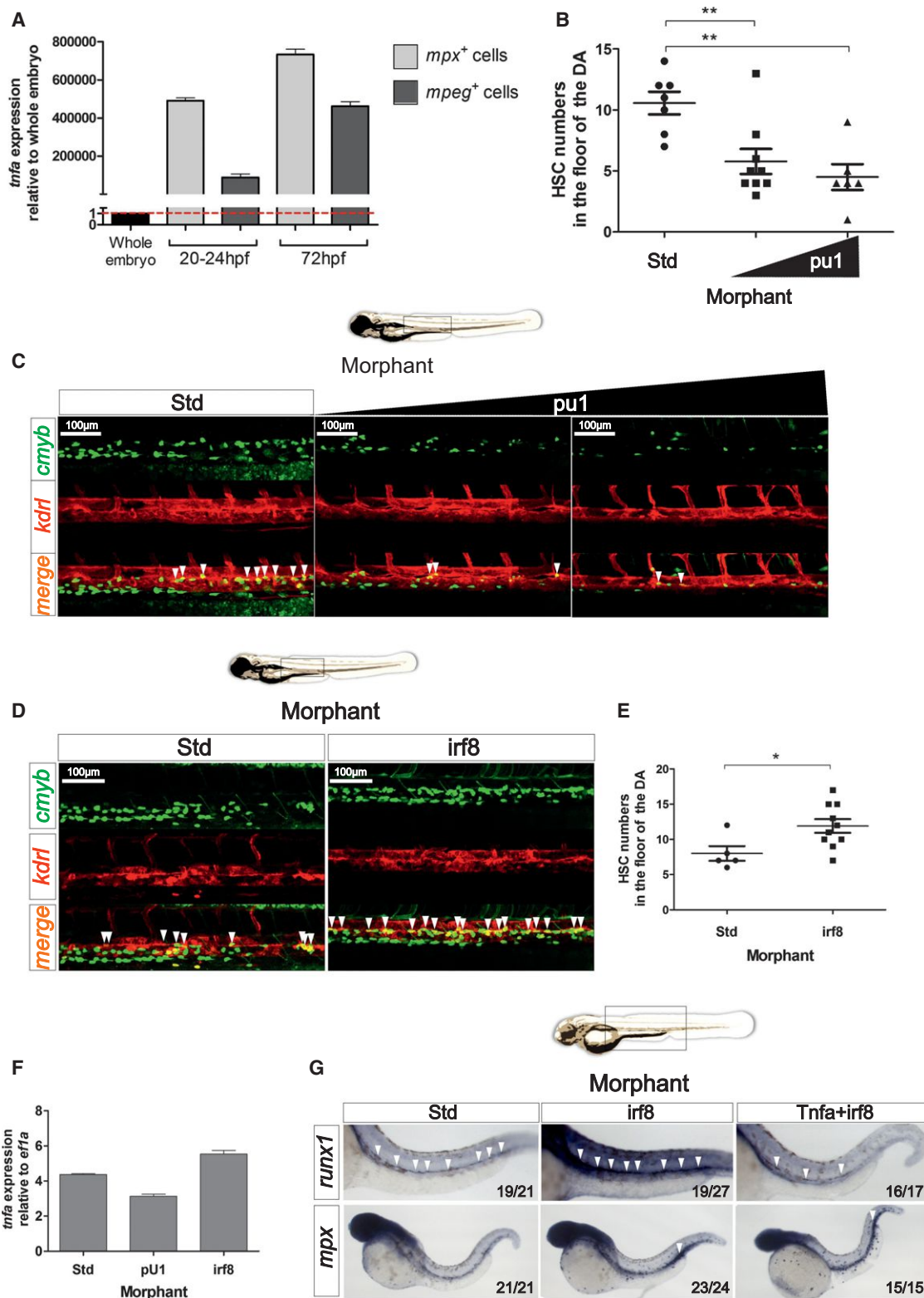


Figure 7. Primitive Myeloid Cells Play a Key Role in HSC Specification

(A) Primitive neutrophils ($mpx:GFP^+$) and macrophages ($mpeg:GFP^+$) were isolated at 20–24 and 72 hpf by FACS and *tnfa* expression was quantified by qPCR. Expression was normalized to *ef1a* and is presented relative to whole-embryo expression. Bars represent means \pm SEM of two independent experiments.

(legend continued on next page)

Detection of Apoptotic Cell Death by TUNEL Labeling

The TUNEL assay was performed as previously described (Espín et al., 2013) with slight modifications. See also Extended Experimental Procedures.

Lipopolysaccharide Injections

Tg(*hsp:Gal4*; *UAS:dn-ikbaa*) embryos were manually dechorionated at 24 hpf, followed by heat shock at 38°C for 50 min. Four hr post-heat-shock, 2 nl of PBS or LPS (900 µg/ml) (L6511, Sigma) was injected into the posterior blood island (PBI). Embryos were then harvested 1 hr postinjection (hpi), and RNA was isolated for qPCR analysis.

Microtome Sections and Immunohistochemistry

Embryos were fixed with 4% PFA, embedded in paraffin, and sectioned at 5 µm in thickness with Leica microtome. Immunohistochemistry was performed as previously described (Kobayashi et al., 2014). The following antibodies were used: mouse anti-mCherry 1:500 (Abcam, ab125096), rabbit anti-p65 (NF-κB) (RB-1638-P; Lab Vision) 1:200, donkey anti-rabbit IgG Alexa Fluor 594-conjugated (Molecular Probes, A-21207) 1:1,000, and donkey anti-mouse IgG Alexa Fluor 488-conjugated (Molecular Probes, A-11029) in addition to DAPI 1:1000 (Life Technologies, D3571).

SUPPLEMENTAL INFORMATION

Supplemental Information includes Extended Experimental Procedures, five figures, one table, and one movie and can be found with this article online at <http://dx.doi.org/10.1016/j.cell.2014.10.031>.

AUTHOR CONTRIBUTIONS

All studies presented herein derive from initial observations by R.E.-P. in the laboratory of V.M. R.E.-P., D.L.S., C.A.C., N.D.C., A.D.K., J.M., D.T., and V.M. designed experiments; R.E.-P., D.L.S., C.A.C., D.G.-M., N.D.C., and S.C. performed research; R.E.-P., D.L.S., C.A.C., D.G.-M., N.D.C., A.D.K., S.C., J.M., V.M., and D.T. analyzed data; and R.E.-P., D.L.S., V.M., and D.T. wrote the paper with minor contributions from the remaining authors.

ACKNOWLEDGMENTS

This work was supported by the Spanish Ministry of Science and Innovation (grants BIO2011-23400, and CSD2007-00002 to V.M.), the Fundación Séneca, Agencia Regional de Ciencia y Tecnología de la Región de Murcia (grant 04538/GERM/06 to V.M. and predoctoral and postdoctoral fellowships to R.E.-P.), the NIH (K01-DK087814-01A1; D.L.S.), a CIRM New Faculty Award RN1-00575-1 (D.T.), AHA Innovative Science Award 12PILT12860010 (D.T.), and NIH R01-DK074482 (D.T.). We thank Inmaculada Fuentes, Pedro Martínez, Efrén Reyes, Chase Melick, and Karen Ong for technical assistance; Nadia Mercader for ideas and discussion; Francisco Juan Martínez-Navarro and Liangdao Li for help with WISH; and Kylie Price and SAI staff for flow cytometry.

Received: April 11, 2014

Revised: September 17, 2014

Accepted: October 15, 2014

Published: November 6, 2014

REFERENCES

- Abbas, S., and Abu-Amer, Y. (2003). Dominant-negative IkappaB facilitates apoptosis of osteoclasts by tumor necrosis factor-alpha. *J. Biol. Chem.* 278, 20077–20082.
- Aggarwal, B.B. (2003). Signalling pathways of the TNF superfamily: a double-edged sword. *Nat. Rev. Immunol.* 3, 745–756.
- Aggarwal, B.B., Gupta, S.C., and Kim, J.H. (2012). Historical perspectives on tumor necrosis factor and its superfamily: 25 years later, a golden journey. *Blood* 119, 651–665.
- Ahn, K.S., and Aggarwal, B.B. (2005). Transcription factor NF-kappaB: a sensor for smoke and stress signals. *Ann. N Y Acad. Sci.* 1056, 218–233.
- Ang, H.L., and Tergaonkar, V. (2007). Notch and NFkappaB signaling pathways: Do they collaborate in normal vertebrate brain development and function? *Bioessays* 29, 1039–1047.
- Baldrige, M.T., King, K.Y., and Goodell, M.A. (2011). Inflammatory signals regulate hematopoietic stem cells. *Trends Immunol.* 32, 57–65.
- Bertrand, J.Y., Kim, A.D., Teng, S., and Traver, D. (2008). CD41+ cmyb+ precursors colonize the zebrafish pronephros by a novel migration route to initiate adult hematopoiesis. *Development* 135, 1853–1862.
- Bertrand, J.Y., Chi, N.C., Santoso, B., Teng, S., Stainier, D.Y., and Traver, D. (2010a). Haematopoietic stem cells derive directly from aortic endothelium during development. *Nature* 464, 108–111.
- Bertrand, J.Y., Cisson, J.L., Stachura, D.L., and Traver, D. (2010b). Notch signaling distinguishes 2 waves of definitive hematopoiesis in the zebrafish embryo. *Blood* 115, 2777–2783.
- Bigas, A., and Espinosa, L. (2012). Hematopoietic stem cells: to be or Notch to be. *Blood* 119, 3226–3235.
- Bigas, A., Robert-Moreno, A., and Espinosa, L. (2010). The Notch pathway in the developing hematopoietic system. *Int. J. Dev. Biol.* 54, 1175–1188.
- Bigas, A., Guiu, J., and Gama-Norton, L. (2013). Notch and Wnt signaling in the emergence of hematopoietic stem cells. *Blood Cells Mol. Dis.* 51, 264–270.
- Boisset, J.C., van Cappellen, W., Andrieu-Soler, C., Galjart, N., Dzierzak, E., and Robin, C. (2010). In vivo imaging of haematopoietic cells emerging from the mouse aortic endothelium. *Nature* 464, 116–120.
- Brown, K.D., Claudio, E., and Siebenlist, U. (2008). The roles of the classical and alternative nuclear factor-kappaB pathways: potential implications for autoimmunity and rheumatoid arthritis. *Arthritis Res. Ther.* 10, 212.
- Burns, C.E., Traver, D., Mayhall, E., Shepard, J.L., and Zon, L.I. (2005). Hematopoietic stem cell fate is established by the Notch-Runx pathway. *Genes Dev.* 19, 2331–2342.
- Cao, Q., Kaur, C., Wu, C.Y., Lu, J., and Ling, E.A. (2011). Nuclear factor-kappa β regulates Notch signaling in production of proinflammatory cytokines and nitric oxide in murine BV-2 microglial cells. *Neuroscience* 192, 140–154.
- Clements, W.K., and Traver, D. (2013). Signalling pathways that control vertebrate haematopoietic stem cell specification. *Nat. Rev. Immunol.* 13, 336–348.
- Cole, L.K., and Ross, L.S. (2001). Apoptosis in the developing zebrafish embryo. *Dev. Biol.* 240, 123–142.

(B) Enumeration of *cmyb*⁺; *kdr1*⁺ HSCs shown in (C). Each dot represents total *cmyb*⁺; *kdr1*⁺ cells per embryo, and black lines indicate means ± SEM for each group of embryos. **p<0.01.

(C) Maximum projections of representative images of *cmyb:GFP*; *kdr1:mCherry* embryos at 48 hpf following injections of Std or pu1 MOs, the latter at two different concentrations. Region shown is the DA, and arrowheads denote *cmyb*⁺; *kdr1*⁺ HSCs.

(D) Maximum projections of representative images of *cmyb:GFP*; *kdr1:mCherry* embryos at 48 hpf in Std and *irf8* morphants. Arrowheads denote *cmyb*⁺; *kdr1*⁺ HSCs.

(E) Enumeration of *cmyb*⁺; *kdr1*⁺ HSCs shown in (D).

(F) *tnfa* expression relative to *ef1a* in 28 hpf Std, pu1, or *irf8* morphants. Bars represent means ± SEM of duplicate samples.

(G) Std, *irf8*, and *irf8+Tnfa* morphants were interrogated by WISH for *runx1* and *mpx* at 28 hpf. All views are lateral, with anteriors to the left. Numbers represent larvae with indicated phenotypes. *p < 0.05 and ***p < 0.001.

See also Figure S5.

- de Bruijn, M.F., Speck, N.A., Peeters, M.C., and Dzierzak, E. (2000). Definitive hematopoietic stem cells first develop within the major arterial regions of the mouse embryo. *EMBO J.* 19, 2465–2474.
- Espín, R., Roca, F.J., Candel, S., Sepulcre, M.P., González-Rosa, J.M., Alcaraz-Pérez, F., Meseguer, J., Cayuela, M.L., Mercader, N., and Mulero, V. (2013). TNF receptors regulate vascular homeostasis in zebrafish through a caspase-8, caspase-2 and P53 apoptotic program that bypasses caspase-3. *Dis. Model. Mech.* 6, 383–396.
- Espinosa, L., Inglés-Esteve, J., Robert-Moreno, A., and Bigas, A. (2003). Ikap-Balpa and p65 regulate the cytoplasmic shuttling of nuclear corepressors: cross-talk between Notch and NF- κ B pathways. *Mol. Biol. Cell* 14, 491–502.
- Espinosa, L., Cathelin, S., D'Altri, T., Trimarchi, T., Statnikov, A., Guiu, J., Rodilla, V., Inglés-Esteve, J., Nomdedeu, J., Bellosillo, B., et al. (2010). The Notch/Hes1 pathway sustains NF- κ B activation through CYLD repression in T cell leukemia. *Cancer Cell* 18, 268–281.
- Faustman, D., and Davis, M. (2010). TNF receptor 2 pathway: drug target for autoimmune diseases. *Nat. Rev. Drug Discov.* 9, 482–493.
- Fernandez, L., Rodriguez, S., Huang, H., Chora, A., Fernandes, J., Mumaw, C., Cruz, E., Pollok, K., Cristina, F., Price, J.E., et al. (2008). Tumor necrosis factor- α and endothelial cells modulate Notch signaling in the bone marrow microenvironment during inflammation. *Exp. Hematol.* 36, 545–558.
- Gering, M., and Patient, R. (2005). Hedgehog signaling is required for adult blood stem cell formation in zebrafish embryos. *Dev. Cell* 8, 389–400.
- Gerondakis, S., Banerjee, A., Grigoriadis, G., Vasanthakumar, A., Gugasyan, R., Sidwell, T., and Grumont, R.J. (2012). NF- κ B subunit specificity in hematopoiesis. *Immunol. Rev.* 246, 272–285.
- Godin, I., and Cumano, A. (2002). The hare and the tortoise: an embryonic hematopoietic race. *Nat. Rev. Immunol.* 2, 593–604.
- Hadland, B.K., Huppert, S.S., Kanungo, J., Xue, Y., Jiang, R., Gridley, T., Conlon, R.A., Cheng, A.M., Kopan, R., and Longmore, G.D. (2004). A requirement for Notch1 distinguishes 2 phases of definitive hematopoiesis during development. *Blood* 104, 3097–3105.
- Herbomel, P., Thisse, B., and Thisse, C. (1999). Ontogeny and behaviour of early macrophages in the zebrafish embryo. *Development* 126, 3735–3745.
- Herbomel, P., Thisse, B., and Thisse, C. (2001). Zebrafish early macrophages colonize cephalic mesenchyme and developing brain, retina, and epidermis through a M-CSF receptor-dependent invasive process. *Dev. Biol.* 238, 274–288.
- Johnston, D.A., Dong, B., and Hughes, C.C. (2009). TNF induction of jagged-1 in endothelial cells is NF- κ B-dependent. *Gene* 435, 36–44.
- Kanther, M., Sun, X., Mühlbauer, M., Mackey, L.C., Flynn, E.J., 3rd, Bagnat, M., Jobin, C., and Rawls, J.F. (2011). Microbial colonization induces dynamic temporal and spatial patterns of NF- κ B activation in the zebrafish digestive tract. *Gastroenterology* 141, 197–207.
- Kim, A.D., Melick, C.H., Clements, W.K., Stachura, D.L., Distel, M., Panáková, D., MacRae, C., Mork, L.A., Crump, J.G., and Traver, D. (2014). Discrete Notch signaling requirements in the specification of hematopoietic stem cells. *EMBO J.* 33, 2363–2373.
- King, K.Y., and Goodell, M.A. (2011). Inflammatory modulation of HSCs: viewing the HSC as a foundation for the immune response. *Nat. Rev. Immunol.* 11, 685–692.
- Kissa, K., and Herbomel, P. (2010). Blood stem cells emerge from aortic endothelium by a novel type of cell transition. *Nature* 464, 112–115.
- Kobayashi, I., Kobayashi-Sun, J., Kim, A.D., Pouget, C., Fujita, N., Suda, T., and Traver, D. (2014). Jam1a-Jam2a interactions regulate haematopoietic stem cell fate through Notch signalling. *Nature* 512, 319–323.
- Kohchi, C., Noguchi, K., Tanabe, Y., Mizuno, D., and Soma, G. (1994). Constitutive expression of TNF- α and - β genes in mouse embryo: roles of cytokines as regulator and effector on development. *Int. J. Biochem.* 26, 111–119.
- Kondo, M., Wagers, A.J., Manz, M.G., Prohaska, S.S., Scherer, D.C., Beilhack, G.F., Shizuru, J.A., and Weissman, I.L. (2003). Biology of hematopoietic stem cells and progenitors: implications for clinical application. *Annu. Rev. Immunol.* 21, 759–806.
- Kopan, R., and Ilgan, M.X. (2009). The canonical Notch signaling pathway: unfolding the activation mechanism. *Cell* 137, 216–233.
- Kortschak, R.D., Tamme, R., and Lardelli, M. (2001). Evolutionary analysis of vertebrate Notch genes. *Dev. Genes Evol.* 211, 350–354.
- Kumano, K., Chiba, S., Kunisato, A., Sata, M., Saito, T., Nakagami-Yamaguchi, E., Yamaguchi, T., Masuda, S., Shimizu, K., Takahashi, T., et al. (2003). Notch1 but not Notch2 is essential for generating hematopoietic stem cells from endothelial cells. *Immunity* 18, 699–711.
- Kyba, M., and Daley, G.Q. (2003). Hematopoiesis from embryonic stem cells: lessons from and for ontogeny. *Exp. Hematol.* 31, 994–1006.
- Lai, E.C. (2004). Notch signaling: control of cell communication and cell fate. *Development* 131, 965–973.
- Langenau, D.M., Ferrando, A.A., Traver, D., Kutok, J.L., Hezel, J.P., Kanki, J.P., Zon, L.I., Look, A.T., and Trede, N.S. (2004). In vivo tracking of T cell development, ablation, and engraftment in transgenic zebrafish. *Proc. Natl. Acad. Sci. USA* 101, 7369–7374.
- Lawson, N.D., Scheer, N., Pham, V.N., Kim, C.H., Chitnis, A.B., Campos-Ortega, J.A., and Weinstein, B.M. (2001). Notch signaling is required for arterial-venous differentiation during embryonic vascular development. *Development* 128, 3675–3683.
- Le Guyader, D., Redd, M.J., Colucci-Guyon, E., Murayama, E., Kissa, K., Briolat, V., Mordelet, E., Zapata, A., Shinomiya, H., and Herbomel, P. (2008). Origins and unconventional behavior of neutrophils in developing zebrafish. *Blood* 111, 132–141.
- Li, L., Jin, H., Xu, J., Shi, Y., and Wen, Z. (2011). Irf8 regulates macrophage versus neutrophil fate during zebrafish primitive myelopoiesis. *Blood* 117, 1359–1369.
- Mizrahi, K., and Askenasy, N. (2014). Physiological functions of TNF family receptor/ligand interactions in hematopoiesis and transplantation. *Blood* 124, 176–183.
- Parsons, M.J., Pisharath, H., Yusuff, S., Moore, J.C., Siekmann, A.F., Lawson, N., and Leach, S.D. (2009). Notch-responsive cells initiate the secondary transition in larval zebrafish pancreas. *Mech. Dev.* 126, 898–912.
- Quillien, A., Moore, J.C., Shin, M., Siekmann, A.F., Smith, T., Pan, L., Moens, C.B., Parsons, M.J., and Lawson, N.D. (2014). Distinct Notch signaling outputs pattern the developing arterial system. *Development* 141, 1544–1552.
- Rhodes, J., Hagen, A., Hsu, K., Deng, M., Liu, T.X., Look, A.T., and Kanki, J.P. (2005). Interplay of pu.1 and gata1 determines myelo-erythroid progenitor cell fate in zebrafish. *Dev. Cell* 8, 97–108.
- Robert-Moreno, A., Guiu, J., Ruiz-Herguido, C., López, M.E., Inglés-Esteve, J., Riera, L., Tipping, A., Enver, T., Dzierzak, E., Gridley, T., et al. (2008). Impaired embryonic haematopoiesis yet normal arterial development in the absence of the Notch ligand Jagged1. *EMBO J.* 27, 1886–1895.
- Roca, F.J., Mulero, I., López-Muñoz, A., Sepulcre, M.P., Renshaw, S.A., Meseguer, J., and Mulero, V. (2008). Evolution of the inflammatory response in vertebrates: fish TNF- α is a powerful activator of endothelial cells but hardly activates phagocytes. *J. Immunol.* 181, 5071–5081.
- Sainson, R.C., Johnston, D.A., Chu, H.C., Holderfield, M.T., Nakatsu, M.N., Crampton, S.P., Davis, J., Conn, E., and Hughes, C.C. (2008). TNF primes endothelial cells for angiogenic sprouting by inducing a tip cell phenotype. *Blood* 111, 4997–5007.
- Santoro, M.M., Samuel, T., Mitchell, T., Reed, J.C., and Stainier, D.Y. (2007). Birc2 (clap1) regulates endothelial cell integrity and blood vessel homeostasis. *Nat. Genet.* 39, 1397–1402.
- Shalaby, M.R., Sundan, A., Loetscher, H., Brockhaus, M., Lesslauer, W., and Espevik, T. (1990). Binding and regulation of cellular functions by monoclonal antibodies against human tumor necrosis factor receptors. *J. Exp. Med.* 172, 1517–1520.
- Shin, H.M., Minter, L.M., Cho, O.H., Gottipati, S., Fauq, A.H., Golde, T.E., Sonenshein, G.E., and Osborne, B.A. (2006). Notch1 augments NF- κ B activity by facilitating its nuclear retention. *EMBO J.* 25, 129–138.

- Song, L.L., Peng, Y., Yun, J., Rizzo, P., Chaturvedi, V., Weijzen, S., Kast, W.M., Stone, P.J., Santos, L., Loreda, A., et al. (2008). Notch-1 associates with IKK α and regulates IKK activity in cervical cancer cells. *Oncogene* 27, 5833–5844.
- Stein, S.J., and Baldwin, A.S. (2013). Deletion of the NF- κ B subunit p50/RelA in the hematopoietic compartment leads to defects in hematopoietic stem cell function. *Blood* 121, 5015–5024.
- Takizawa, H., Boettcher, S., and Manz, M.G. (2012). Demand-adapted regulation of early hematopoiesis in infection and inflammation. *Blood* 119, 2991–3002.
- Thisse, C., Thisse, B., Schilling, T.F., and Postlethwait, J.H. (1993). Structure of the zebrafish *snail1* gene and its expression in wild-type, spadetail and no tail mutant embryos. *Development* 119, 1203–1215.
- van der Vaart, M., van Soest, J.J., Spaik, H.P., and Meijer, A.H. (2013). Functional analysis of a zebrafish *myd88* mutant identifies key transcriptional components of the innate immune system. *Dis. Model. Mech.* 6, 841–854.
- Wang, H., Tian, Y., Wang, J., Phillips, K.L., Binch, A.L., Dunn, S., Cross, A., Chiverton, N., Zheng, Z., Shapiro, I.M., et al. (2013). Inflammatory cytokines induce NOTCH signaling in nucleus pulposus cells: implications in intervertebral disc degeneration. *J. Biol. Chem.* 288, 16761–16774.
- Westerfield, M. (2000). *The Zebrafish Book. A guide for the laboratory use of zebrafish (Danio rerio)*, Fourth Edition (Eugene: Univ. of Oregon Press).
- Wiens, G.D., and Glenney, G.W. (2011). Origin and evolution of TNF and TNF receptor superfamilies. *Dev. Comp. Immunol.* 35, 1324–1335.
- Wynn, T.A., Chawla, A., and Pollard, J.W. (2013). Macrophage biology in development, homeostasis and disease. *Nature* 496, 445–455.
- Zhao, C., Xiu, Y., Ashton, J., Xing, L., Morita, Y., Jordan, C.T., and Boyce, B.F. (2012). Noncanonical NF- κ B signaling regulates hematopoietic stem cell self-renewal and microenvironment interactions. *Stem Cells* 30, 709–718.
- Zhao, J.L., Ma, C., O'Connell, R.M., Mehta, A., DiLoreto, R., Heath, J.R., and Baltimore, D. (2014). Conversion of danger signals into cytokine signals by hematopoietic stem and progenitor cells for regulation of stress-induced hematopoiesis. *Cell Stem Cell* 14, 445–459.

RACK1 Controls IRES-Mediated Translation of Viruses

Karim Majzoub,¹ Mohamed Lamine Hafirassou,^{2,3} Carine Meignin,^{1,2} Akira Goto,¹ Stefano Marzi,⁴ Antonina Fedorova,^{2,3} Yann Verdier,⁵ Joëlle Vinh,⁵ Jules A. Hoffmann,^{1,2,6} Franck Martin,⁴ Thomas F. Baumert,^{2,3,7,8} Catherine Schuster,^{2,3,8,*} and Jean-Luc Imler^{1,2,8,*}

¹CNRS UPR9022, Institut de Biologie Moléculaire et Cellulaire, 67000 Strasbourg, France

²Université de Strasbourg, 67000 Strasbourg, France

³Inserm UMR1110, Institut de Recherche sur les Maladies Virales et Hépatiques, 67000 Strasbourg, France

⁴CNRS UPR9002, Institut de Biologie Moléculaire et Cellulaire, 67000 Strasbourg, France

⁵USR3149, ESPCI ParisTech, 75005 Paris, France

⁶Institut d'Etudes Avancées de l'Université de Strasbourg, 67000 Strasbourg, France

⁷Institut Hospitalo-Universitaire (IHU), Pôle hépato-digestif, Hôpitaux Universitaires de Strasbourg, 67000 Strasbourg, France

⁸Co-senior author

*Correspondence: catherine.schuster@unistra.fr (C.S.), jl.imler@unistra.fr (J.-L.I.)

<http://dx.doi.org/10.1016/j.cell.2014.10.041>

SUMMARY

Fighting viral infections is hampered by the scarcity of viral targets and their variability, resulting in development of resistance. Viruses depend on cellular molecules—which are attractive alternative targets—for their life cycle, provided that they are dispensable for normal cell functions. Using the model organism *Drosophila melanogaster*, we identify the ribosomal protein RACK1 as a cellular factor required for infection by internal ribosome entry site (IRES)-containing viruses. We further show that RACK1 is an essential determinant for hepatitis C virus translation and infection, indicating that its function is conserved for distantly related human and fly viruses. Inhibition of RACK1 does not affect *Drosophila* or human cell viability and proliferation, and RACK1-silenced adult flies are viable, indicating that this protein is not essential for general translation. Our findings demonstrate a specific function for RACK1 in selective mRNA translation and uncover a target for the development of broad antiviral intervention.

INTRODUCTION

Viral infections are a significant threat for all living organisms. In humans, acute and chronic viral infections cause a wide spectrum of diseases, including life-threatening inflammation and cancer. A major challenge for the control of viral infections is that viruses, due to the small size of their genomes, offer few intrinsic targets either for recognition by the immune system or for inhibition by antiviral effector molecules. Furthermore, the error-prone viral polymerases allow RNA viruses to rapidly escape detection by the immune system and to resist the adverse effects of directly acting antiviral molecules. Significantly, viruses rely on numerous host factors for essential functions during their life cycle.

These are not subject to rapid sequence changes and hence provide good alternative targets for antiviral therapy. Therefore, a central challenge is to identify cellular factors required for viral replication but dispensable for normal cell function.

RNA replication, transcription, and translation are critical steps in the life cycle of RNA viruses that involve interactions with host-cell molecules. In the model organism *Drosophila melanogaster*, the small interfering RNA (siRNA) pathway targets viral RNAs (reviewed in Ding [2010]). In order to better characterize the contribution of the three core components of this pathway, Dicer-2, R2D2, and AGO2, we performed a proteomic analysis of the complexes assembling around these molecules in infected *Drosophila* cells (K.M., J.-L.I., Y.V., and J.V., unpublished data). One protein copurifying with R2D2 and AGO2 in cells infected with the picorna-like *Drosophila* C virus (DCV) was the evolutionarily conserved ribosomal protein RACK1. The RACK1 protein has been extensively studied during the last two decades and has been shown to be involved in different aspects of cell regulation. RACK1 is an adaptor protein, interacting with a variety of signaling molecules (e.g., PKC, Src, and MAPK) (Belozherov et al., 2014; Gibson, 2012; Long et al., 2014) and is a component of the 40S subunit of the ribosome (Coyle et al., 2009; Sengupta et al., 2004). RACK1 is thus ideally suited to connect signal transduction pathways to the regulation of translation (Nilsson et al., 2004). Indeed, RACK1 was found to interact with the initiation factor eIF6, which associates with the 60S subunit of the ribosome and prevents its association with the 40S subunit. eIF6 phosphorylation by RACK1-assisted PKC triggers its release from the 60S subunit, thus promoting the formation of 80S active ribosomes (Ceci et al., 2003).

Here, we show that RACK1 is mandatory for DCV replication but is largely dispensable for cell viability and proliferation. We further demonstrate that RACK1 is required for internal ribosome entry site (IRES)-dependent translation in *Drosophila* and in human hepatocytes, where this factor is an essential determinant of hepatitis C virus infection. By contrast, RACK1 is not required for 5' cap-dependent translation. Collectively, our data unravel a specific function for ribosomal protein RACK1

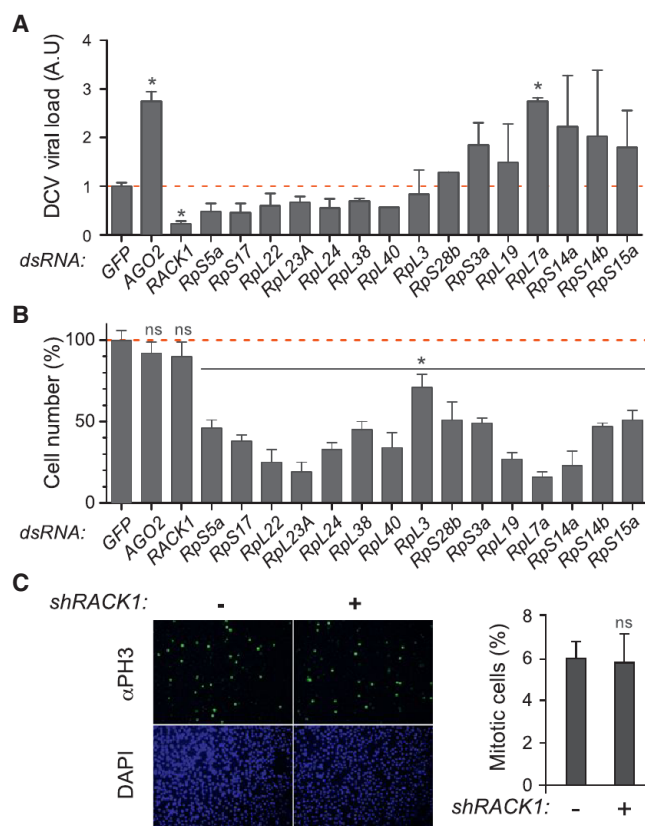


Figure 1. RACK1 Is Required for DCV Replication, but Not for Viability or Proliferation in *Drosophila* Cells

(A and B) Quantification of DCV viral RNA levels by qRT-PCR (A) and of cell numbers as estimated by DAPI staining (B) in cells treated with the indicated dsRNAs to induce silencing. Cells treated with a dsRNA corresponding to GFP and AGO2 sequences are used as a reference and a control, respectively.

(C) S2 cells stably transfected with a metallothionein promoter-driven vector expressing a shRNA targeting the 5' UTR from the *RACK1* gene were treated or not with CuSO₄ for 3 days, stained with DAPI and an anti-phospho-H3 antibody (left), and counted (right).

Data represent the mean and SEM of at least three independent experiments. ns, nonsignificant; *p < 0.05. See also Figure S1 and Table S1.

in selective mRNA translation of fly and human viruses and uncover a target for the development of broad antiviral intervention.

RESULTS

RACK1 Is Required for *Dicistroviridae* Infection in *Drosophila*

In a proteomic analysis of the interactome of Dicer-2, R2D2, and AGO2 in virus-infected cells (K.M., J.-L.I., Y.V., and J.V., unpublished data), we identified 16 ribosomal proteins. To address the functional relevance of this finding, we systematically depleted these ribosomal proteins from S2 cells by RNAi and tested DCV replication. Knockdown of most ribosomal genes affected cell viability or proliferation and did not yield interpretable results regarding DCV infection (Figures 1A and 1B). Indeed, silencing of these genes may result in decreased ability of the cells either to

support viral replication or to control the infection. By contrast, depletion of RACK1 (Figure S1A available online) did not affect cell viability or proliferation in S2 cells (Figures 1B and 1C) or in two other cell lines (Figure S1B). However, it resulted in a significant decrease of DCV titer in infected cells (Figure 1A). Furthermore, RACK1 silencing did not affect replication of either flock house virus (FHV) or vesicular stomatitis virus (VSV) (Figures 2A and 2B), indicating that the RACK1-depleted cells are not only viable and able to proliferate but can also support replication of other viruses. To test whether the effect of RACK1 was specific to DCV or to the family to which it belongs, we infected S2 cells with Cricket Paralysis Virus (CrPV), another member of the *Dicistroviridae* family. Replication of CrPV was also strongly impaired when RACK1 was depleted (Figure 2B).

We next confirmed these findings in vivo. RACK1 null mutant flies are not viable, indicating that RACK1 exerts developmental functions (Kadmas et al., 2007). In agreement with this finding, silencing RACK1 expression with a small hairpin RNA (shRNA) driven by the broadly active *actin5C* promoter was embryonic lethal. When the thermosensitive Gal80 system was used to express the shRNA only in adult flies, development occurred normally and the adult flies expressed significantly reduced levels of RACK1 at the permissive temperature of 29°C (Figure 2C). The reduced levels of RACK1 did not affect the viability of the flies, although it reduced longevity by 20% at this temperature. In addition, the eggs laid by RACK1-silenced females showed a phenotype similar to that of RACK1 mutants (Figure S1C) (Kadmas et al., 2007). Thus, even though RACK1 is required during development, it appears to be largely dispensable in adult flies. As expected, when these flies were challenged with DCV, both viral RNA and capsid protein levels were markedly reduced at 1 and 2 days postinfection compared to controls (Figure 2D). Overall, our data indicate that replication of the *Dicistroviridae* DCV and CrPV requires the ribosomal factor RACK1, which is otherwise dispensable for the viability of S2 cells and adult flies.

RACK1 Is Required for Viral IRES-Dependent Translation

Our data indicate that RACK1 is required for a step of viral replication specific to *Dicistroviridae*. Whereas FHV and VSV use a canonical strategy of cap-dependent initiation of translation, DCV and CrPV RNA recruits the 40S ribosomal subunit through IRES sequences to initiate translation (Figure S2A). Furthermore, although initially identified as a scaffolding protein involved in protein kinase C signaling, RACK1 is now recognized as a component of the 40S subunit of the ribosome. This suggested to us that RACK1 was required for viral translation. We first verified that RACK1 is indeed required at the ribosome level for CrPV replication. We silenced RACK1 expression in a stable cell line using an shRNA targeting the 5' UTR (Figure S1D) and observed a marked decrease in CrPV replication (Figure S1E). Transfection of a vector expressing wild-type RACK1 restored CrPV replication in these cells (Figure 3A). By contrast, expression of mutant proteins unable to interact with either RpS17 (D108Y) (Kuroha et al., 2010) or 18S rRNA (R38D/K40E) (Coyle et al., 2009) did not rescue CrPV replication (Figure 3A). We conclude that RACK1 is required in the 40S ribosomal subunit for CrPV replication.

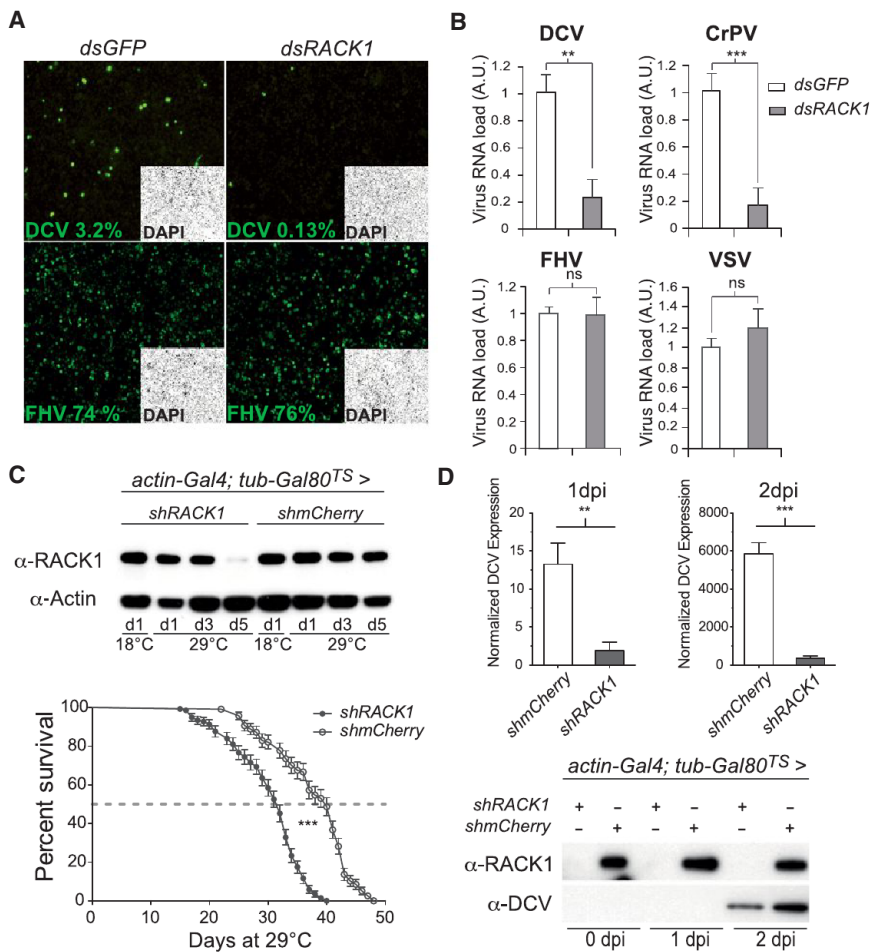


Figure 2. RACK1 Is Required for Replication of DCV and CrPV, but Not FHV and VSV

(A and B) S2 cells were treated with either control (GFP) or RACK1 dsRNA for 4 days before challenge with DCV, FHV, VSV, or CrPV. Viral infection was monitored by immunofluorescence using antibodies recognizing capsid proteins (A) and qRT-PCR (B) after 16 hr or, in the case of VSV, after 48 hr. The percentage of infected cells is indicated for each virus in (A).

(C) Silencing of RACK1 expression in transgenic flies expressing a shRNA targeting the 5' UTR from the RACK1 gene using the Gal4-UAS system and the broadly expressed actin-Gal4 driver controlled by the thermosensitive (TS) tub-Gal80 repressor. A shRNA targeting the mCherry protein was used as a control. The lifespan of RACK1-depleted flies is shown in the bottom graph.

(D) RACK1-silenced flies infected by DCV after 5 days at 29°C show a decrease of the viral RNA and protein, as indicated by qRT-PCR (top) and western blot.

Data represent the mean and SEM from at least three independent experiments. ns, nonsignificant; dpi, days postinfection; *p < 0.05, **p < 0.01, and ***p < 0.001. See also Figures S1 and S2.

To confirm that RACK1 is involved in translation from *Dicistroviridae* RNAs, we tested whether its depletion affected translation of luciferase reporters placed under the control of the two IRES elements from CrPV (Figures 3B and S2C). Translation of a 5' cap-dependent RNA was not affected in the absence of RACK1, although it was affected when expression of eIF4E was knocked down. Translation from the CrPV 5' IRES reporter was not reduced, and was even slightly increased, when eIF4E was silenced, suggesting that the 5' IRES drives noncanonical translation. Interestingly, a significant reduction of luciferase production was observed for the 5' IRES reporter in RACK1-silenced cells (Figure 3B). Silencing of RACK1 did not affect the amount of the 5' IRES reporter luciferase mRNA in the cells, indicating that RACK1 affects translation, rather than RNA stability (Figure S3). By contrast, translation driven by the intergenic (IGR) IRES (Jan and Sarnow, 2002; Spahn et al., 2004) was not affected by the level of RACK1 in the cells (Figure 3B). Polysome profiles from S2 cells and RACK1-silenced stable derivatives of these cells (Figure S1D) were similar, confirming that RACK1 does not affect significantly general translation (Figure 3C). Finally, we prepared cell-free translation extracts from control and RACK1-depleted S2 cells and used them to monitor translation of in-vitro-transcribed, capped, and IRES-

dependent RNAs. Translation of the 5' IRES reporter RNA was strongly reduced in the RACK1 depleted extract. By contrast, translation of the 5' CAP and IGR IRES-dependent reporters was not inhibited and was even slightly stimulated (Figure 3D). Overall, our data indicate that ribosomal RACK1 is required

RACK1 Is an Essential Host Factor for HCV Infection

RACK1 is an evolutionarily strongly conserved factor, and we asked whether it plays a role in the translation driven by the IRES of a mammalian virus. Hepatitis C virus (HCV), a major cause of liver disease and hepatocellular carcinoma, is a positive strand RNA virus member of the *Flaviviridae* family depending on a highly structured IRES for its translation (Figure S2) (Spahn et al., 2001). Transfection of an siRNA targeting RACK1 markedly reduced expression of the protein in Huh7.5.1 cells (Figure 4A), a human hepatocyte-derived cell line highly permissive for HCV infection (Lindenbach et al., 2005; Wakita et al., 2005). Infection of RACK1-depleted Huh7.5.1 cells by cell-culture-derived HCV (Jc1 strain) was strongly and significantly reduced, as revealed both by immunodetection of the viral core protein (Figure 4A) and the focus-forming assay performed by infection of naive Huh7.5.1 cells with supernatants from infected and treated cells (Figure 4B). A similar inhibition of infection was observed for HCV Luc-Jc1 (Figure 4C), a well-characterized recombinant virus expressing a luciferase reporter (Figure S2B). Inhibition of RACK1 expression was as efficient as the silencing of the key HCV

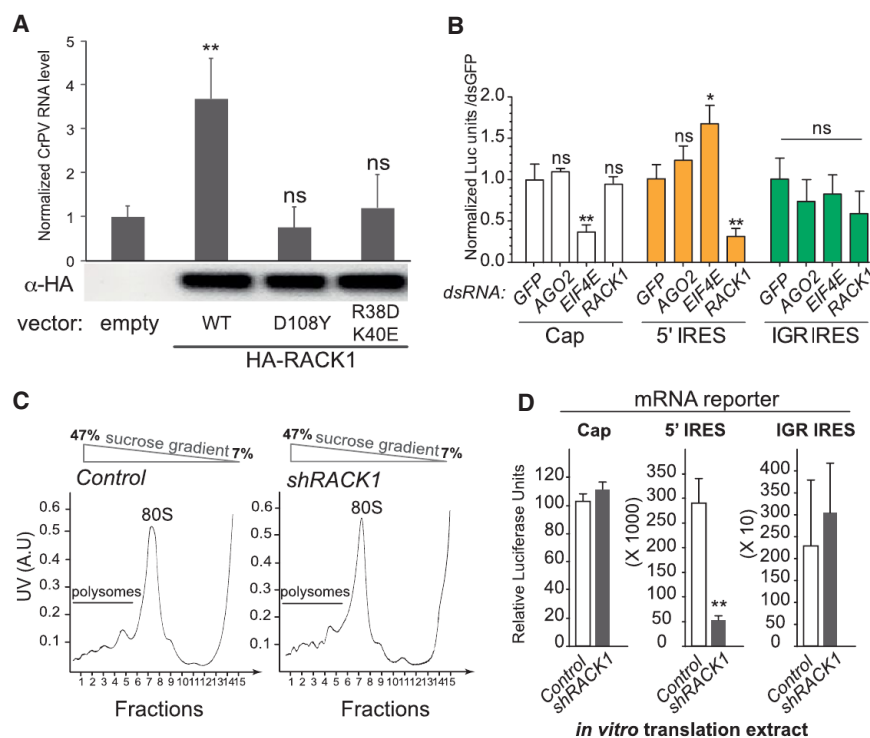


Figure 3. The Ribosomal Protein RACK1 Is Required for IRES-Mediated Translation

(A) Stable S2 transformants expressing a shRNA targeting the 5' UTR of *RACK1* were transfected with vectors expressing three versions of *RACK1* (WT, D108Y, or R38D/K40E). Expression of the transfected *RACK1* was monitored by western blot using an antibody recognizing the N-terminal tag HA. The cells were infected with CrPV for 16 hr, and viral RNA loads were determined by qRT-PCR. Data represent the mean and SEM from three independent experiments.

(B) *RACK1* is required for translation regulated by the 5' IRES, but not the IGR IRES, of CrPV. S2 cells were treated with dsRNAs corresponding to *GFP* (control), *AGO2*, *eIF4E*, or *RACK1* for 3 days before transfection of the indicated Luciferase reporters (5' CAP, IRES_{CrPV-IGR}, or IRES_{CrPV-5'}; see Figure S2). Luciferase activity was monitored 48 hr later. The ratio of the activity of the IRES-dependent luciferase and the 5' cap-dependent luciferase is plotted and normalized to the control for the three reporters. Data represent the mean and SD from six independent experiments.

(C) Polysome profiles from S2 cells expressing or not expressing a shRNA targeting the 5' UTR of *RACK1*. The position of the peaks corresponding to the 80S ribosomes and the polysomes are indicated.

(D) In vitro translation of capped and IRES-dependent reporters using cell-free extracts prepared from control or *RACK1*-silenced S2 cells. Data represent the mean and SD from three independent experiments. ns, nonsignificant; **p* < 0.05 and ***p* < 0.01. See also Figures S2 and S3.

host factors CD81 (Koutsoudakis et al., 2007) and Cyclophilin A (CypA) (Kaul et al., 2009) (Figures 4A–4C). We next transiently depleted *RACK1* in Huh7.5.1 cells replicating the reporter virus HCV Luc-Jc1 and observed a marked impairment of HCV replication (Figure 4D), demonstrating that *RACK1* is required for HCV translation/replication rather than entry. HCV replication rebound observed after day 4 was due to progressive loss of *RACK1* silencing, leading to neosynthesis of *RACK1* (Figure S4A).

To confirm that the inhibition of HCV replication is indeed mediated by the effect of *RACK1* on IRES-mediated translation, we established stable cell lines expressing an IRES_{HCV}-luciferase reporter construct or a classical capped reporter gene (Figure S2C) and transfected these cells with *RACK1*-specific siRNAs. Silencing of *RACK1* markedly and specifically decreased IRES_{HCV}-dependent translation to a similar extent as an antiviral siRNA directed against the IRES_{HCV} (Figure 4E). By contrast, silencing of ribosomal protein RPS3 inhibited translation from both IRES- and 5' cap-dependent reporter constructs (Figure 4E). Similar results were obtained when in-vitro-transcribed reporter mRNAs were transfected into Huh7.5.1 cells, ruling out an effect of *RACK1* on transcription of the IRES_{HCV}-luciferase reporter gene (Figure 4F).

Importantly, *RACK1*-specific siRNAs did not affect cell proliferation (Figure S4B) or viability, in contrast to silencing of the ribosomal protein RPS3 (Figure 4G). A genome-wide microarray analysis of polysomes prepared from control or *RACK1*-silenced human Huh7.5.1 cells revealed that the amount in polysomes of mRNAs for most genes, including house-keeping genes and important hepatocyte-specific genes such as albumin or lipopro-

teins, was not affected by *RACK1* depletion (Figure 4H). Of note, silencing of *RACK1* also did not affect the presence of 5' terminal oligopyrimidine tract (TOP) mRNAs in polysomes (for details, see Supplemental Information). This result suggests that translation of the large majority of mRNAs is not affected by the absence of *RACK1* in human hepatocytes under normal culture conditions and confirms the results obtained in the model organism *Drosophila*.

The Effect of *RACK1* on Viral Translation Is Independent of the miRNA Pathway

While this work was in progress, a role for *RACK1* in miRNA function was reported in the plant *Arabidopsis thaliana* (Speth et al., 2013), the model organism *Caenorhabditis elegans* (Chu et al., 2014; Jannot et al., 2011), and humans (Otsuka et al., 2011). In light of the important impact of the cellular microRNA miR122 on HCV replication (Jopling et al., 2005), this suggested that *RACK1* might operate on viral translation through the miRNA pathway. We first verified that *RACK1* affects the miRNA pathway in *Drosophila*. Expression in S2 cells of two previously described miRNA reporters, *Par-6* and *nerfin-1* (Eulalio et al., 2007), was derepressed when *RACK1* was silenced, indicating that, in *Drosophila* as well, *RACK1* is involved in miRNA function (Figures 5A and 5B). We note, however, that the derepression is much stronger for the miR1 reporter than for the miR9b reporter, suggesting that the role of *RACK1* may be specific of a subset of miRNAs. By contrast, silencing of *Dcr-1* or *AGO1* derepressed equally well the two miR reporters (Figures 5A and 5B). We next tested whether miRNAs play a role in viral replication by

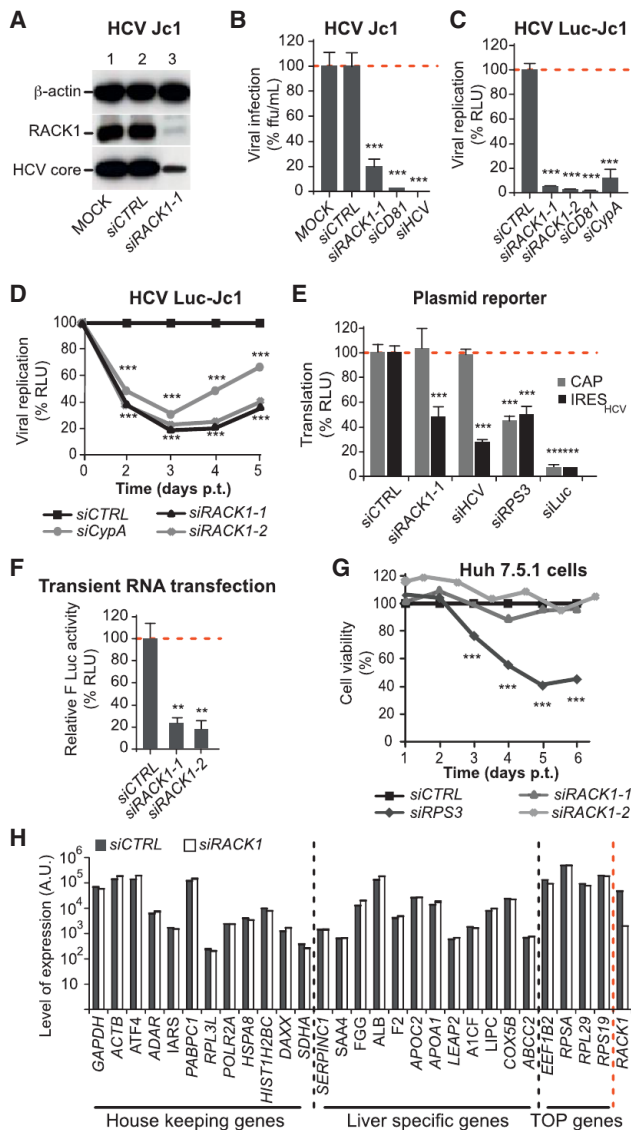


Figure 4. RACK1 Is a Specific Host Factor Required for IRES-Mediated Translation of HCV

(A–C) Huh7.5.1 cells were transfected with siRNAs either control (siCTRL) or targeting RACK1 (siRACK1-1 or -2), CD81 (siCD81), Cyclophilin A (siCypA), or HCV IRES (siHCV) before infection 3 days later with HCV Jc1 (A and B) or HCV Luc-Jc1 (C). Viral infection was monitored 3 days postinfection by immunoblotting using antibodies recognizing HCV core protein (A), by counting foci forming units (ffu/mL) (B), or by quantifying luciferase activity (C).

(D) HCV Luc-Jc1 replicating cells were transfected with siCTRL and two different siRNAs targeting RACK1 or siCypA, and replication was monitored during 5 days by luciferase activity quantification.

(E) Huh7.5.1 cell lines stably expressing an IRES (IRES_{HCV}-Luc) or a 5' cap (CTRL-Luc)-dependent luciferase reporter gene were transfected with siCTRL, siRACK1, siHCV, siRPS3, or siLuc. Translation was monitored 72 hr later by luciferase activity quantification.

(F) Huh7.5.1 cells were transfected with the indicated siRNAs and 72 hr later with in-vitro-transcribed IRES_{HCV} or 5' cap-dependent luciferase mRNAs. Luciferase activity was monitored 5 hr later.

(G) Cell viability of Huh7.5.1 cells silenced with the indicated siRNAs was measured during 5 days using MTT assay. **p < 0.01 and ***p < 0.005. For (B)–(G), data represent the mean ± SD of at least three experiments.

monitoring accumulation of viral RNAs in cells depleted of *Dcr-1* or *AGO1*. Silencing of *Dcr-1* had no effect on the viral RNA load of the four viruses tested (Figure 5C). Silencing of *AGO1* did reduce, to some extent, CrPV and DCV RNA load. However, this reduction was variable in the case of DCV and not to the extent of the reduction observed when *RACK1* was silenced for DCV and CrPV (Figure 5C). Thus, although the miRNA pathway may have a contribution in the replication of *Dicistroviridae*, our data suggest that the strong effect of *RACK1* cannot be accounted for only by its effect on miRNA function. This was confirmed by the observation that silencing of *Dcr-1* or *AGO1* had no effect on translation driven by the IRES_{CrPV-5'}, unlike silencing of *RACK1* (Figure 5D).

In mammalian hepatocytes, HCV translation depends on *AGO2* and *miR122* (Conrad et al., 2013; Roberts et al., 2011). As expected, transfection of Huh7.5.1 cells with a *miR122* mimic increased HCV replication, whereas transfection of a *miR122* inhibitor led to decreased viral replication (Figure 5E). Importantly, the impact of the *miR122* mimic and the *miR122* inhibitor on HCV replication did not depend on *RACK1* (Figure 5E). To unambiguously determine whether the contribution of *RACK1* to HCV translation was dependent on *miR122*, we used HEK293T cells, which do not express *miR122* (Da Costa et al., 2012; Figure S5). Silencing of *RACK1* expression efficiently repressed translation driven by the IRES_{HCV} in these cells (Figure 5F). Finally, transduction of HEK293T cells with an expression vector for *miR122* did not affect the impact of *RACK1* on HCV translation (Figure 5G), although *miR122* was expressed and functional in these cells (Figures S5A and S5B). Collectively, these results indicate that *RACK1* and *miR122* regulate HCV translation by different mechanisms.

The eIF3j Subunit Is Dispensable for Cell Viability but Is Important for CrPV and HCV Replication

We next attempted to gain mechanistic insight on the role of *RACK1* in viral translation. Previous cryo-electron microscopy studies have highlighted the interaction of the 40S subunit with the HCV IRES and, in spite of their low resolution, have suggested that binding of the HCV IRES triggers a pronounced conformational change in the small subunit of the ribosome (Spahn et al., 2001, 2004). HCV IRES has been also visualized on the 80S human ribosome and *RACK1* localized in its vicinity (Boehringer et al., 2005; Sengupta et al., 2004). The recently elucidated crystal structure of the small subunit of the ribosome at 3.9 Å (Rabl et al., 2011) allows us to fit the crystal structure in the cryo-electron microscopy density. The picture obtained suggests that *RACK1* is located in close proximity to the IRES of HCV in the region affected by the conformational change triggered upon IRES_{HCV} binding (Figure S6A). By contrast, the IRES_{CrPV-IGR}, which does not depend on *RACK1* (Figure 2C), interacts with a distinct site of the 40S subunit, directly contacting

(H) Quantification of representative mRNAs in polysomes prepared from Huh7.5.1 cells transfected with siCTRL or siRACK1. Gene expression levels, shown in arbitrary units, were determined by hybridization on genome-wide microarrays and represent the mean ± SD of four individual samples. Each sample was analyzed individually. See also Figures S2 and S4 and Table S2.

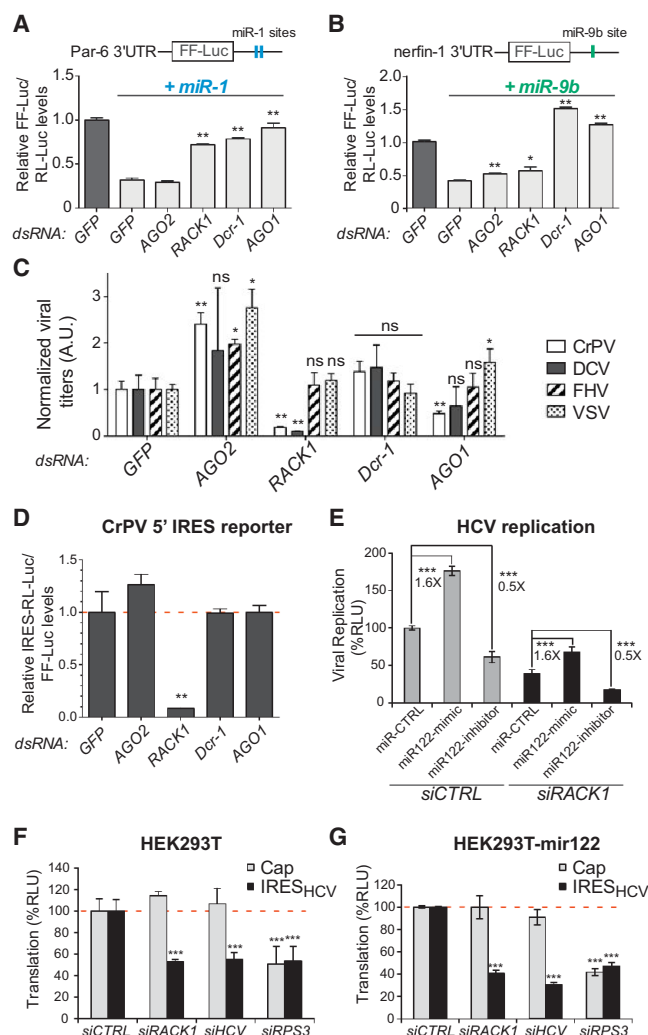


Figure 5. The Effect of RACK1 on Viral Translation Is Independent of the miRNA Pathway

(A and B) RACK1 is required for miR1 and miR9b silencing. The structure of the *Par-6* 3' UTR and *nerfin-1* 3' UTR reporter constructs is represented on top, and the luciferase activity in cells silenced for the indicated genes is shown below.

(C) Effect of the depletion of *AGO1*, *Dcr-1*, and *RACK1* on replication in *Drosophila* S2 cells of CrPV, DCV, FHV, and VSV. Cells were transfected with the indicated dsRNAs and infected 4 days later. Viral RNA was extracted 24 hpi and was quantified by qRT-PCR.

(D) Silencing of *AGO1* or *Dcr-1* does not affect the activity of a Luciferase reporter gene controlled by the IRES_{CrPV-5'} in *Drosophila* S2 cells.

(E) A miR122 mimic and a miR122 inhibitor affect HCV replication similarly in control or *RACK1*-silenced Huh7.5.1 hepatocytes.

(F and G) Silencing of *RACK1* affects the activity of the IRES_{HCV}-luciferase reporter in miR122-deficient (F) and stably transfected miR122 expressing (G) HEK293T cells, respectively.

Data represent the mean and SEM of at least three independent experiments. ns, nonsignificant; **p* < 0.05, ***p* < 0.01, and ****p* < 0.001. See also Figure S5.

RpS25 (Figure S6B) (Fernández et al., 2014; Koh et al., 2014; Schüller et al., 2006; Spahn et al., 2004). Although no direct contacts between RACK1 and IRES_{HCV} could be observed, a recent study indicates that a peripheral domain of the translation initia-

tion factor eIF3, which is required for IRES_{HCV}-dependent translation (Kieft, 2008), is in contact with RACK1 (Figure S6C) (Hashem et al., 2013a; Sun et al., 2013). This domain may be the functional link between RACK1 and IRES_{HCV}-dependent translation.

We asked whether some subunits of eIF3, such as eIF3c, which has been shown to interact with RACK1 in yeast, may be specifically involved in IRES-dependent translation, like RACK1. We first tested in *Drosophila* S2 cells whether some subunits of the eIF3 complex are dispensable for cell viability in normal culture conditions. Out of the 14 genes encoding eIF3 components (the *Drosophila* genome contains two *eIF3g* paralogues, *CG8636* (*eIF3ga*) and *CG10881* (*eIF3gb*)), only two were not required for cell viability or proliferation (Figure 6A). One of these genes is *CG10881*, encoding eIF3gb, which is expressed specifically in testis (Chintapalli et al., 2007) and thus provides a useful negative control. The second gene is the *Drosophila* ortholog of *eIF3j* (Figure 6A). We next monitored CrPV replication in cells silenced for eIF3j or eIF3gb (*CG10881*). Although silencing of eIF3gb did not affect CrPV replication, silencing of eIF3j resulted in a significant reduction of CrPV replication (Figure 6B). Silencing of eIF3j, but not of *eIF3gb*, also affected translation of the IRES_{CrPV5'}-luciferase reporter, although not as strongly as silencing of RACK1 (Figure 6C). In Huh7.5.1 cells, silencing of eIF3c affected cell viability. By contrast, silencing of eIF3j only marginally affected cell viability (Figure 6D; Wagner et al., 2014). Interestingly, however, it resulted in a moderate but significant decrease of HCV replication (Figure 6E). Altogether, these results suggest that the eIF3j subunit might participate in the observed effects of RACK1 on translation.

DISCUSSION

A Function for RACK1 in IRES-Dependent Translation

Our data reveal a function for RACK1 in specific mRNA translation. Indeed, silencing RACK1 expression does not affect viability of *Drosophila* S2 or human Huh7.5.1 cells in tissue culture, indicating that formation of active ribosomes is not strictly dependent on RACK1. In vivo as well, translation can occur in the absence of RACK1, as lethality in *RACK1* mutant animals does not occur before larval stages for *Drosophila* and gastrulation in mice (Kadmas et al., 2007; Volta et al., 2013). In agreement with this observation, translation of a 5' cap-dependent reporter was not affected in the absence of RACK1 in *Drosophila* and human cells. Nevertheless, the fact that *RACK1* mutant animals cannot complete their development suggests that this protein is required for the translation of some cellular mRNAs, in addition to viral IRES-containing RNAs. Interestingly, previous studies have highlighted the role of another protein from the 40S subunit of the ribosome, Rps25, in IRES-dependent translation (Landry et al., 2009). Performed on yeast and mammalian tissue-culture cells with IRES reporter assays, these experiments concluded that Rps25 is essential for the activity of two viral IRES, IRES_{HCV} and IRES_{CrPV-IGR}. The mechanism used by Rps25 and RACK1 to promote translation is probably different because (1) Rps25 is required for IRES_{CrPV-IGR}, unlike RACK1, and (2) structural data place Rps25 at a distance from RACK1

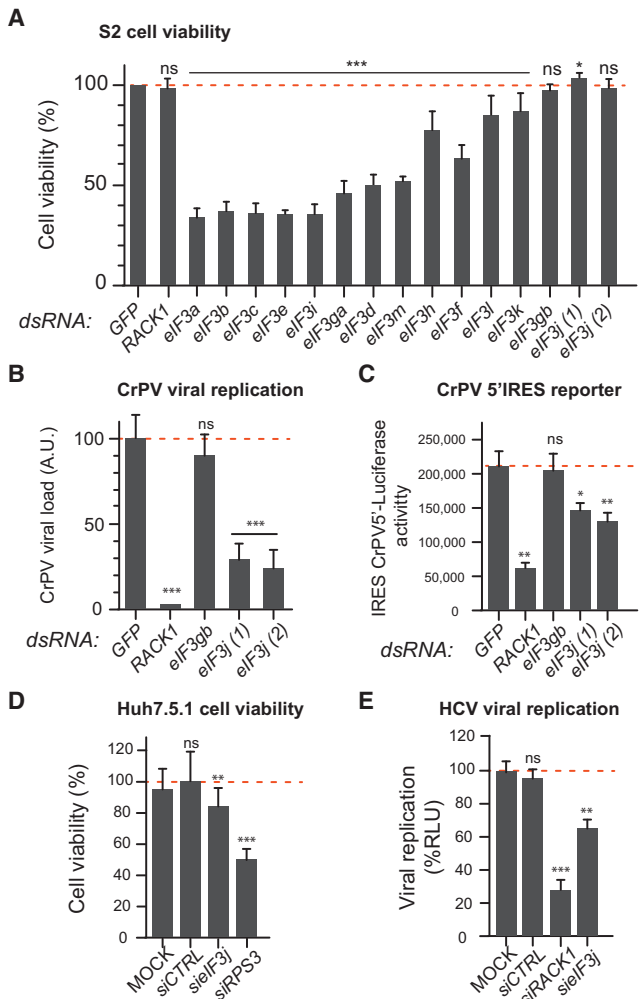


Figure 6. eIF3j Is Required for CrPV and HCV Replication, but Not for Cell Viability

(A) Quantification by the MTS assay of the number of viable cells 5 days after treatment of S2 cells with the indicated dsRNAs. Two different dsRNA preparations, targeting different regions of the gene, were used for *eIF3j*.

(B) Quantification by qRT-PCR of CrPV viral RNA levels in S2 cells treated with the indicated dsRNAs.

(C) Activity of the IRES_{CrPV5'} in S2 cells silenced for the indicated genes.

(D) Quantification of Huh7.5.1 cell viability after silencing of the indicated genes.

(E) Quantification of HCV replication in Huh7.5.1 cells transfected with the indicated siRNAs.

Data represent the mean and SD of at least three experiments. See also Figure S6.

on the 40S subunit of the ribosome, providing an explanation for its importance on the activity of the IRES_{CrPV-IGR}. Several other ribosomal proteins (e.g., Rpl38 and Rpl40) were recently proposed to be involved in specific translation of some 5' cap-dependent mRNAs (Kondrashov et al., 2011; Lee et al., 2013), indicating that transcript-specific regulation can occur in the absence of IRES elements. Our data lend support to an evolving picture of the eukaryotic ribosome, which includes structurally peripheral components such as RACK1 involved in the modula-

tion of translation of specific mRNAs (reviewed in Xue and Barna [2012]). They have implications for the development of new antivirals and raise questions on the mechanism underlying the role of RACK1 in IRES-dependent translation.

RACK1 as a Target for Broad Antiviral Intervention

Our results open interesting therapeutic perspectives for a broad range of viral infections, including chronic hepatitis C, a major cause of liver cirrhosis and cancer. Because HCV translation initiates viral genome neosynthesis via the formation of the replication complex, RACK1-mediated translation is a crucial step in virus propagation. Thus, RACK1 is a novel host target for antiviral therapy, which is complementary to interferon-based therapies or direct-acting antivirals (DAAs). DAAs have achieved high response rates with cure in late-stage clinical trials, but high costs will limit their broad access. In addition, certain patient groups (e.g., genotype 3, renal failure, hepatic decompensation, and liver transplantation) will need complementary approaches (Chung and Baumert, 2014; Liang and Ghany, 2013).

The low variability of host factors targeted by host-targeted antivirals (HTAs) results in a high genetic barrier to resistance (Nathan, 2012). Indeed, HTAs effectively inhibit HCV escape variants (Fofana et al., 2010; Lupberger et al., 2011), as well as DAA-resistant virus (Xiao et al., 2014a). Furthermore, their complementary mechanism of action results in synergy with DAAs (Xiao et al., 2014b). Given that HTAs interfere with host targets, one theoretical caveat is the possibly greater risk of cellular toxicity as compared to DAAs. Interestingly, our data obtained in cell culture models did not reveal any major toxicity linked to RACK1 inhibition. Thus, our proof-of-concept studies in state-of-the-art cell culture models open a highly attractive and innovative perspective to develop small molecules targeting RACK1. RACK1 inhibitors may also be of interest for treatment of infection of many other human or animal viruses using 5' cap-independent mechanisms for the translation of their RNAs.

Mechanistic Insight on the Role of RACK1 in IRES-Dependent Translation

While this work was in preparation, several reports described a role for RACK1 in miRNA function. However, our data in *Drosophila* and human cells indicate that the role of RACK1 in IRES-dependent translation does not involve small regulatory RNAs. Nevertheless, the connection between RACK1 and AGO proteins is intriguing and suggests that RACK1 may participate in a checkpoint for the control of the translation of specific mRNAs by miRNAs or siRNAs.

The ribosome code or filter hypothesis posits that some ribosomal proteins have evolved to mediate translation of specific mRNAs (Mauro and Edelman, 2002; Topisirovic and Sonenberg, 2011; Xue and Barna, 2012). A central unresolved issue of this hypothesis is the nature of the *cis*-acting elements defining a possible "ribosome code." In the case of RACK1, these *cis*-acting elements include viral IRES. Interestingly, the IRES_{CrPV-IGR} is active in the absence of RACK1, unlike the IRES_{CrPV-5'} or the IRES_{HCV}. This IRES_{CrPV-IGR} (class I IRES) is capable on its own, without any initiation factors, of binding directly the 40S subunit and of recruiting the 60S subunit to form an active 80S ribosome, thus bypassing the loading of the initiator methionyl-tRNAi

(Jan and Sarnow, 2002; Pestova et al., 2004). By contrast, the function of IRES_{HCV} (class II IRES) requires two canonical eIFs, eIF2 and 3, as well as Met-tRNAi (Kieft, 2008). This suggests that the effect of RACK1 on translation initiation may require one of these factors. Interestingly, the eIF3 complex binds to the 40S ribosomal subunit and to the IRES_{HCV} (e.g., Kieft et al., 2001). Furthermore, RACK1 was shown to associate with one of the eIF3 subunits in order to assemble a translation preinitiation complex in yeast (Hashem et al., 2013a; Kouba et al., 2012).

Although our understanding of the molecular structure of the core of the 13 subunits eIF3 complex has progressed remarkably in recent years (e.g., Hashem et al., 2013b; Sun et al., 2011), the role of the noncore subunits remains essentially untested in animals. Interestingly, the subunit eIF3e in the yeast *Schizosaccharomyces pombe* is involved in translation of a selected set of RNAs (Sha et al., 2009; Zhou et al., 2005). More recently, one of the two *eIF3h* genes present in zebrafish, *eIF3ha*, was shown to encode a factor specifically targeting crystalline isoform mRNAs for translation during lens development (Choudhuri et al., 2013). Our data indicate that, like RACK1, the subunit eIF3j is not required for cell viability in *Drosophila* but is required for CrPV replication and IRES_{CrPV5'}-driven translation. This raises the possibility that RACK1 and eIF3j act together in translation of a specific subset of mRNAs.

Several observations support a role for eIF3j in selective mRNA translation. First, it is located in the decoding center of the 40S ribosomal subunit, where it can regulate access to the mRNA binding cleft (Fraser et al., 2007; 2009). Second, it is located at the periphery of the eIF3 complex, often in substoichiometric quantities, indicating that it can undergo regulated cycles of association and dissociation (Hinnebusch, 2006; Miyamoto et al., 2005; Sha et al., 2009). Third, experiments in *S. pombe* and human cells indicate that it can be regulated posttranslationally by phosphorylation (Sha et al., 2009) or caspase-mediated C-terminal truncation (Bushell et al., 2000). Altogether, this suggests that RACK1 may act as a scaffold recruiting an enzyme modifying eIF3j in order to allow access of the entry channel of the 40S subunit to IRES-containing mRNAs. In a way, such a scenario would be reminiscent of the recently described role of another eIF3 subunit, eIF3e, which controls the recruitment of the kinase Mnk1 to phosphorylate eIF4E, thus promoting selective mRNA translation in human cells (Walsh and Mohr, 2014).

EXPERIMENTAL PROCEDURES

Silencing Candidate Gene Expression by RNAi and Screening

dsRNAs targeting the candidate genes were designed using the E-RNAi algorithm (<http://www.dkfz.de/signaling/e-mai3/>) (Table S1). Knockdown in *Drosophila* S2 cells was performed in 96-well plates using the bathing method, and cells were challenged with virus 4 days later. Viral load was determined by qRT-PCR. Alternatively, infected cells were fixed and labeled with anti-capsid antibodies for immunofluorescence analysis using the InCELL1000 Analyzer workstation (GE LifeSciences). Image data processing was performed using the InCELL Analyzer software. See Extended Experimental Procedures in Supplemental Information for more details.

Preparation of Cell-free Extract for In Vitro Translation

In vitro translation competent extracts were prepared from control or RACK1-silenced S2 cells as described in Wakiyama et al., (2006). Briefly, cells were resuspended in lysis solution (40 mM HEPES-KOH [pH 8], 100 mM potassium

acetate, 1 mM magnesium acetate, and 1 mM dithiothreitol) at a cell density of $\sim 10^9$ ml⁻¹ and were placed in the Cell Disruption Bomb (Parr Instrument Company). The homogenate produced upon the pressure release was cleared by centrifugations at 4°C, and creatine kinase was added at 0.24 mg.ml⁻¹ of lysate before storage in aliquotes at -80°C. Reporter mRNAs were synthesized by transcription in vitro using recombinant T7 RNA polymerase. A nonfunctional cap (ApppG) (New England Biolabs) was added at the 5' end of the IRES monocistronic reporter mRNAs to protect them from degradation. Cap-dependent translation was measured with a Renilla Luciferase reporter mRNA that was capped with the ScriptCap m7G capping system (Epicenter Biotechnologies). In vitro translation was performed as previously described (Wakiyama et al., 2006) and under subsaturating conditions to avoid substrate titration.

HCV Infection and Replication Assays

Huh7.5.1 human hepatoma cells were infected with cell-culture-derived HCV (HCVcc strains Jc1 and Luc-Jc1, half-maximal tissue culture infectious dose [TCID50 10⁴ ml⁻¹ for both viruses]) as described (Lupberger et al., 2011; Pietschmann et al., 2006). Two days before infection, gene silencing was performed by reverse transfection with 10 nM of siRNA (Silencer Select siRNA, Ambion) specific for RACK1, CD81, Cyclophilin A, HCV IRES, or a nonspecific control siRNA. Viral infection and RACK1 depletion were analyzed by western blotting and quantified by counting of focus forming units (ffu)/ml following immunostaining using a HCV core-specific antibody (mAbC7-50, Affinity Bio-Reagents, CO) or by luciferase reporter gene expression in cell lysates 3 days postinfection. For HCV replication experiments, Huh7.5.1 cells were electroporated with HCV Luc-Jc1 RNA (Koutsoudakis et al., 2007). Cells were reverse transfected with siRNAs 3 days later. Global analysis of polysome-associated mRNAs in control and RACK1 silenced Huh7.5.1 cells was performed on Agilent "SurePrint G3 Human Gene Expression v2" 8x60k microarray (Table S2 and see Extended Experimental Procedures for more details).

ACCESSION NUMBERS

The microarray data presented in Figure 4H have been deposited in the Gene Expression Omnibus database under the accession number GSE60374.

SUPPLEMENTAL INFORMATION

Supplemental Information includes Extended Experimental Procedures, six figures, and two tables and can be found with this article online at <http://dx.doi.org/10.1016/j.cell.2014.10.041>.

AUTHOR CONTRIBUTIONS

K.M. and M.L.H. performed and analyzed most experiments in *Drosophila* and Huh7.5.1 cells, respectively. C.M. and A.G. designed, performed, or supervised and analyzed the data for the experiments in flies and in Figure 6. S.M. prepared Figure S6, and A.F. analyzed the polysome microarray expression data. Y.V. and J.V. identified RACK1 as a protein associating with AGO2 and R2D2 in DCV-infected cells. F.M. designed, performed, or supervised and analyzed the data for the polysome and in vitro translation experiments. J.A.H., T.F.B., C.S., and J.-L.I. designed the experiments, analyzed the data, and wrote the article.

ACKNOWLEDGMENTS

This work was supported by CNRS, Inserm, and grants from NIH (PO1 A1070167 to J.A.H. and J.-L.I.); Investissement d'Avenir programs (HepSys ANR-10-LAB-28 to T.B. and NetRNA ANR-10-LABX-36; I2MC ANR-11-EQPX-0022, to J.-L.I. and S.M.; IDEX W13RATCS to C.S.); ANR (ANR-09-MIEN-006-01 to J.-L.I. and ANR-11-SVSE802501 to F.M.); ARC (IHU201301187 to T.B.); ANRS (2010-307/2011-415, to C.S.); and FRM (to C.M.). Financial support from the TGE FT-ICR, the ESPCI ParisTech, and the Fonds pour la Recherche et la Technologie of the French Ministry of Research

for the LTQ FT acquisition are gratefully acknowledged. M.L.H., K.M., and A.F. were supported by fellowships from ANRS, CNRS/Région Alsace, and the IDEX program from the University of Strasbourg, respectively. We thank S. Pellegrino and M. Yusupov (IGBMC, Strasbourg) for advice and assistance in the preparation of translation competent cell free extracts; R. Bartenschlager (University of Heidelberg, Heidelberg, Germany) for providing Jc1 and Luc-Jc1 expression vectors; F.V. Chisari (TSRI, La Jolla) for Huh7.5.1 cells; M. Beckerle (University of Utah) for providing RACK1 antiserum; S. Thompson (Birmingham) for sharing the plasmid pSRT208; and the transgenic RNAi Project (TriP) at Harvard Medical School (NIH/NIGMS R01-GM084947) for providing transgenic RNAi fly stocks. We also thank E. Santiago, N. Ftaich, and M. Parrot for technical assistance; B. Fischer and D. Dembele (siRNA screening and microarray platforms at IGBMC) for the InCell1000 and the transcriptomic analysis; and E. Westhof and W. Filipowicz for discussions and critical reading of the manuscript.

Received: August 13, 2014

Revised: September 16, 2014

Accepted: October 20, 2014

Published: November 20, 2014

REFERENCES

- Belozerov, V.E., Ratkovic, S., McNeill, H., Hilliker, A.J., and McDermott, J.C. (2014). In vivo interaction proteomics reveal a novel p38 mitogen-activated protein kinase/RACK1 pathway regulating proteostasis in *Drosophila* muscle. *Mol. Cell. Biol.* **34**, 474–484.
- Boehringer, D., Thermann, R., Ostareck-Lederer, A., Lewis, J.D., and Stark, H. (2005). Structure of the hepatitis C virus IRES bound to the human 80S ribosome: remodeling of the HCV IRES. *Structure* **13**, 1695–1706.
- Bushell, M., Wood, W., Clemens, M.J., and Morley, S.J. (2000). Changes in integrity and association of eukaryotic protein synthesis initiation factors during apoptosis. *Eur. J. Biochem.* **267**, 1083–1091.
- Ceci, M., Gaviraghi, C., Gorrini, C., Sala, L.A., Offenhäuser, N., Marchisio, P.C., and Biffo, S. (2003). Release of eIF6 (p27BBP) from the 60S subunit allows 80S ribosome assembly. *Nature* **426**, 579–584.
- Chintapalli, V.R., Wang, J., and Dow, J.A.T. (2007). Using FlyAtlas to identify better *Drosophila melanogaster* models of human disease. *Nat. Genet.* **39**, 715–720.
- Choudhuri, A., Maitra, U., and Evans, T. (2013). Translation initiation factor eIF3h targets specific transcripts to polysomes during embryogenesis. *Proc. Natl. Acad. Sci. USA* **110**, 9818–9823.
- Chu, Y.-D., Wang, W.-C., Chen, S.-A.A., Hsu, Y.-T., Yeh, M.-W., Slack, F.J., and Chan, S.-P. (2014). RACK-1 regulates let-7 microRNA expression and terminal cell differentiation in *Caenorhabditis elegans*. *Cell Cycle* **13**, 1995–2009.
- Chung, R.T., and Baumert, T.F. (2014). Curing chronic hepatitis C—the arc of a medical triumph. *N. Engl. J. Med.* **370**, 1576–1578.
- Conrad, K.D., Giering, F., Erfurth, C., Neumann, A., Fehr, C., Meister, G., and Niepmann, M. (2013). MicroRNA-122 dependent binding of Ago2 protein to hepatitis C virus RNA is associated with enhanced RNA stability and translation stimulation. *PLoS ONE* **8**, e56272.
- Coyle, S.M., Gilbert, W.V., and Doudna, J.A. (2009). Direct link between RACK1 function and localization at the ribosome in vivo. *Mol. Cell. Biol.* **29**, 1626–1634.
- Da Costa, D., Turek, M., Felmlee, D.J.D., Girardi, E., Pfeffer, S., Long, G., Bartenschlager, R., Zeisel, M.B.M., and Baumert, T.F.T. (2012). Reconstitution of the entire hepatitis C virus life cycle in nonhepatic cells. *J. Virol.* **86**, 11919–11925.
- Ding, S.W. (2010). RNA-based antiviral immunity. *Nat. Rev. Immunol.* **10**, 632–644.
- Eulalio, A., Rehwinkel, J., Stricker, M., Huntzinger, E., Yang, S.-F., Doerks, T., Dörner, S., Bork, P., Boutros, M., and Izaurralde, E. (2007). Target-specific requirements for enhancers of decapping in miRNA-mediated gene silencing. *Genes Dev.* **21**, 2558–2570.
- Fernández, I.S., Bai, X.-C., Murshudov, G., Scheres, S.H.W., and Ramakrishnan, V. (2014). Initiation of translation by cricket paralysis virus IRES requires its translocation in the ribosome. *Cell* **157**, 823–831.
- Fofana, I., Krieger, S.E., Grunert, F., Glauben, S., Xiao, F., Fafi-Kremer, S., Soulier, E., Royer, C., Thumann, C., Mee, C.J., et al. (2010). Monoclonal anti-claudin 1 antibodies prevent hepatitis C virus infection of primary human hepatocytes. *Gastroenterology* **139**, 953–964, e1–e4.
- Fraser, C.S., Berry, K.E., Hershey, J.W.B., and Doudna, J.A. (2007). eIF3j is located in the decoding center of the human 40S ribosomal subunit. *Mol. Cell* **26**, 811–819.
- Fraser, C.S., Hershey, J.W.B., and Doudna, J.A. (2009). The pathway of hepatitis C virus mRNA recruitment to the human ribosome. *Nat. Struct. Mol. Biol.* **16**, 397–404.
- Gibson, T.J. (2012). RACK1 research - ships passing in the night? *FEBS Lett.* **586**, 2787–2789.
- Hashem, Y., des Georges, A., Dhote, V., Langlois, R., Liao, H.Y., Grassucci, R.A., Hellen, C.U., Pestova, T.V., and Frank, J. (2013a). Structure of the mammalian ribosomal 43S preinitiation complex bound to the scanning factor DHX29. *Cell* **153**, 1108–1119.
- Hashem, Y., des Georges, A., Dhote, V., Langlois, R., Liao, H.Y., Grassucci, R.A., Pestova, T.V., Hellen, C.U., and Frank, J. (2013b). Hepatitis-C-virus-like internal ribosome entry sites displace eIF3 to gain access to the 40S subunit. *Nature* **503**, 539–543.
- Hinnebusch, A.G. (2006). eIF3: a versatile scaffold for translation initiation complexes. *Trends Biochem. Sci.* **31**, 553–562.
- Jan, E., and Sarnow, P. (2002). Factorless ribosome assembly on the internal ribosome entry site of cricket paralysis virus. *J. Mol. Biol.* **324**, 889–902.
- Jannot, G., Bajan, S., Giguère, N.J., Bouasker, S., Banville, I.H., Piquet, S., Hutvagner, G., and Simard, M.J. (2011). The ribosomal protein RACK1 is required for microRNA function in both *C. elegans* and humans. *EMBO Rep.* **12**, 581–586.
- Jopling, C.L., Yi, M., Lancaster, A.M., Lemon, S.M., and Sarnow, P. (2005). Modulation of hepatitis C virus RNA abundance by a liver-specific MicroRNA. *Science* **309**, 1577–1581.
- Kadmas, J.L., Smith, M.A., Pronovost, S.M., and Beckerle, M.C. (2007). Characterization of RACK1 function in *Drosophila* development. *Dev. Dyn.* **236**, 2207–2215.
- Kaul, A., Stauffer, S., Berger, C., Pertel, T., Schmitt, J., Kallis, S., Zayas, M., Lohmann, V., Luban, J., and Bartenschlager, R. (2009). Essential role of cyclophilin A for hepatitis C virus replication and virus production and possible link to polyprotein cleavage kinetics. *PLoS Pathog.* **5**, e1000546.
- Kieft, J.S. (2008). Viral IRES RNA structures and ribosome interactions. *Trends Biochem. Sci.* **33**, 274–283.
- Kieft, J.S., Zhou, K., Jubin, R., and Doudna, J.A. (2001). Mechanism of ribosome recruitment by hepatitis C IRES RNA. *RNA* **7**, 194–206.
- Koh, C.S., Briot, A.F., Grigorieff, N., and Korostelev, A.A. (2014). Taura syndrome virus IRES initiates translation by binding its tRNA-mRNA-like structural element in the ribosomal decoding center. *Proc. Natl. Acad. Sci. USA* **111**, 9139–9144.
- Kondrashov, N., Pusic, A., Stumpf, C.R., Shimizu, K., Hsieh, A.C., Xue, S., Ishijima, J., Shiroishi, T., and Barna, M. (2011). Ribosome-mediated specificity in Hox mRNA translation and vertebrate tissue patterning. *Cell* **145**, 383–397.
- Kouba, T., Rutkai, E., Karásková, M., and Valášek, L. (2012). The eIF3c/NIP1 PCI domain interacts with RNA and RACK1/ASC1 and promotes assembly of translation preinitiation complexes. *Nucleic Acids Res.* **40**, 2683–2699.
- Koutsoudakis, G., Herrmann, E., Kallis, S., Bartenschlager, R., and Pietschmann, T. (2007). The level of CD81 cell surface expression is a key determinant for productive entry of hepatitis C virus into host cells. *J. Virol.* **81**, 588–598.
- Kuroha, K., Akamatsu, M., Dimitrova, L., Ito, T., Kato, Y., Shirahige, K., and Inada, T. (2010). Receptor for activated C kinase 1 stimulates nascent polypeptide-dependent translation arrest. *EMBO Rep.* **11**, 956–961.

- Landry, D.M., Hertz, M.I., and Thompson, S.R. (2009). RPS25 is essential for translation initiation by the Dicistroviridae and hepatitis C viral IRESs. *Genes Dev.* 23, 2753–2764.
- Lee, A.S.-Y., Burdeinick-Kerr, R., and Whelan, S.P.J. (2013). A ribosome-specialized translation initiation pathway is required for cap-dependent translation of vesicular stomatitis virus mRNAs. *Proc. Natl. Acad. Sci. USA* 110, 324–329.
- Liang, T.J., and Ghany, M.G. (2013). Current and future therapies for hepatitis C virus infection. *N. Engl. J. Med.* 368, 1907–1917.
- Lindenbach, B.D., Evans, M.J., Syder, A.J., Wölk, B., Tellinghuisen, T.L., Liu, C.C., Maruyama, T., Hynes, R.O., Burton, D.R., McKeating, J.A., and Rice, C.M. (2005). Complete replication of hepatitis C virus in cell culture. *Science* 309, 623–626.
- Long, L., Deng, Y., Yao, F., Guan, D., Feng, Y., Jiang, H., Li, X., Hu, P., Lu, X., Wang, H., et al. (2014). Recruitment of phosphatase PP2A by RACK1 adaptor protein deactivates transcription factor IRF3 and limits type I interferon signaling. *Immunity* 40, 515–529.
- Lupberger, J., Zeisel, M.B., Xiao, F., Thumann, C., Fofana, I., Zona, L., Davis, C., Mee, C.J., Turek, M., Gorke, S., et al. (2011). EGFR and EphA2 are host factors for hepatitis C virus entry and possible targets for antiviral therapy. *Nat. Med.* 17, 589–595.
- Mauro, V.P., and Edelman, G.M. (2002). The ribosome filter hypothesis. *Proc. Natl. Acad. Sci. USA* 99, 12031–12036.
- Miyamoto, S., Patel, P., and Hershey, J.W. (2005). Changes in ribosomal binding activity of eIF3 correlate with increased translation rates during activation of T lymphocytes. *J. Biol. Chem.* 280, 28251–28264.
- Nathan, C. (2012). Fresh approaches to anti-infective therapies. *Sci. Transl. Med.* 4, 140sr2.
- Nilsson, J., Sengupta, J., Frank, J., and Nissen, P. (2004). Regulation of eukaryotic translation by the RACK1 protein: a platform for signalling molecules on the ribosome. *EMBO Rep.* 5, 1137–1141.
- Otsuka, M., Takata, A., Yoshikawa, T., Kojima, K., Kishikawa, T., Shibata, C., Takekawa, M., Yoshida, H., Omata, M., and Koike, K. (2011). Receptor for activated protein kinase C: requirement for efficient microRNA function and reduced expression in hepatocellular carcinoma. *PLoS ONE* 6, e24359.
- Pestova, T.V., Lomakin, I.B., and Hellen, C.U. (2004). Position of the CrPV IRES on the 40S subunit and factor dependence of IRES/80S ribosome assembly. *EMBO Rep.* 5, 906–913.
- Pietschmann, T., Kaul, A., Koutsoudakis, G., Shavinskaya, A., Kallis, S., Steinmann, E., Abid, K., Negro, F., Dreux, M., Cosset, F.L., and Bartenschlager, R. (2006). Construction and characterization of infectious intragenotypic and intergenotypic hepatitis C virus chimeras. *Proc. Natl. Acad. Sci. USA* 103, 7408–7413.
- Rabl, J., Leibundgut, M., Ataide, S.F., Haag, A., and Ban, N. (2011). Crystal structure of the eukaryotic 40S ribosomal subunit in complex with initiation factor 1. *Science* 331, 730–736.
- Roberts, A.P.E., Lewis, A.P., and Jopling, C.L. (2011). miR-122 activates hepatitis C virus translation by a specialized mechanism requiring particular RNA components. *Nucleic Acids Res.* 39, 7716–7729.
- Schüler, M., Connell, S.R., Lescoute, A., Giesebrecht, J., Dabrowski, M., Schroer, B., Mielke, T., Penczek, P.A., Westhof, E., and Spahn, C.M. (2006). Structure of the ribosome-bound cricket paralysis virus IRES RNA. *Nat. Struct. Mol. Biol.* 13, 1092–1096.
- Sengupta, J., Nilsson, J., Gursky, R., Spahn, C.M., Nissen, P., and Frank, J. (2004). Identification of the versatile scaffold protein RACK1 on the eukaryotic ribosome by cryo-EM. *Nat. Struct. Mol. Biol.* 11, 957–962.
- Sha, Z., Brill, L.M., Cabrera, R., Kleefeld, O., Scheliga, J.S., Glickman, M.H., Chang, E.C., and Wolf, D.A. (2009). The eIF3 interactome reveals the translatome, a supercomplex linking protein synthesis and degradation machineries. *Mol. Cell* 36, 141–152.
- Spahn, C.M., Kieft, J.S., Grassucci, R.A., Penczek, P.A., Zhou, K., Doudna, J.A., and Frank, J. (2001). Hepatitis C virus IRES RNA-induced changes in the conformation of the 40S ribosomal subunit. *Science* 291, 1959–1962.
- Spahn, C.M., Jan, E., Mulder, A., Grassucci, R.A., Sarnow, P., and Frank, J. (2004). Cryo-EM visualization of a viral internal ribosome entry site bound to human ribosomes: the IRES functions as an RNA-based translation factor. *Cell* 118, 465–475.
- Speth, C., Willing, E.-M., Rausch, S., Schneeberger, K., and Laubinger, S. (2013). RACK1 scaffold proteins influence miRNA abundance in Arabidopsis. *Plant J.* 76, 433–445.
- Sun, C., Todorovic, A., Querol-Audí, J., Bai, Y., Villa, N., Snyder, M., Ashchyan, J., Lewis, C.S., Hartland, A., Gradia, S., et al. (2011). Functional reconstitution of human eukaryotic translation initiation factor 3 (eIF3). *Proc. Natl. Acad. Sci. USA* 108, 20473–20478.
- Sun, C., Querol-Audí, J., Mortimer, S.A., Arias-Palomo, E., Doudna, J.A., Nogales, E., and Cate, J.H.D. (2013). Two RNA-binding motifs in eIF3 direct HCV IRES-dependent translation. *Nucleic Acids Res.* 41, 7512–7521.
- Topisirovic, I., and Sonenberg, N. (2011). Translational control by the eukaryotic ribosome. *Cell* 145, 333–334.
- Volta, V., Beugnet, A., Gallo, S., Magri, L., Brina, D., Pesce, E., Calamita, P., Sanvito, F., and Biffo, S. (2013). RACK1 depletion in a mouse model causes lethality, pigmentation deficits and reduction in protein synthesis efficiency. *Cell. Mol. Life Sci.* 70, 1439–1450.
- Wagner, S., Herrmannová, A., Malík, R., Peclínová, L., and Valášek, L.S. (2014). Functional and biochemical characterization of human eukaryotic translation initiation factor 3 in living cells. *Mol. Cell. Biol.* 34, 3041–3052.
- Wakita, T., Pietschmann, T., Kato, T., Date, T., Miyamoto, M., Zhao, Z., Murthy, K., Habermann, A., Kräusslich, H.G., Mizokami, M., et al. (2005). Production of infectious hepatitis C virus in tissue culture from a cloned viral genome. *Nat. Med.* 11, 791–796.
- Wakiyama, M., Kaitsu, Y., and Yokoyama, S. (2006). Cell-free translation system from *Drosophila* S2 cells that recapitulates RNAi. *Biochem. Biophys. Res. Commun.* 343, 1067–1071.
- Walsh, D., and Mohr, I. (2014). Coupling 40S ribosome recruitment to modification of a cap-binding initiation factor by eIF3 subunit e. *Genes Dev.* 28, 835–840.
- Xiao, F., Fofana, I., Heydmann, L., Barth, H., Soulier, E., Habersetzer, F., Doffoël, M., Bukh, J., Patel, A.H., Zeisel, M.B., and Baumert, T.F. (2014a). Hepatitis C virus cell-cell transmission and resistance to direct-acting antiviral agents. *PLoS Pathog.* 10, e1004128.
- Xiao, F., Fofana, I., Thumann, C., Mailly, L., Alles, R., Robinet, E., Meyer, N., Schaeffer, M., Habersetzer, F., Doffoël, M., et al. (2014b). Synergy of entry inhibitors with direct-acting antivirals uncovers novel combinations for prevention and treatment of hepatitis C. *Gut*. Published online May 21, 2014. <http://dx.doi.org/10.1136/gutjnl-2013-306155>.
- Xue, S., and Barna, M. (2012). Specialized ribosomes: a new frontier in gene regulation and organismal biology. *Nat. Rev. Mol. Cell Biol.* 13, 355–369.
- Zhou, C., Arslan, F., Wee, S., Krishnan, S., Ivanov, A.R., Oliva, A., Leatherwood, J., and Wolf, D.A. (2005). PCI proteins eIF3e and eIF3m define distinct translation initiation factor 3 complexes. *BMC Biol.* 3, 14.

Global Changes in the RNA Binding Specificity of HIV-1 Gag Regulate Virion Genesis

Sebla B. Kutluay,¹ Trinity Zang,^{1,2} Daniel Blanco-Melo,¹ Chelsea Powell,¹ David Jannain,¹ Manel Errando,³ and Paul D. Bieniasz^{1,2,*}

¹Laboratory of Retrovirology

²Howard Hughes Medical Institute

Aaron Diamond AIDS Research Center, The Rockefeller University, 455 First Avenue, New York, NY 10016, USA

³Department of Physics and Astronomy, Barnard College, Columbia University, New York, NY 10027, USA

*Correspondence: pbienias@adarc.org

<http://dx.doi.org/10.1016/j.cell.2014.09.057>

SUMMARY

The HIV-1 Gag protein orchestrates all steps of virion genesis, including membrane targeting and RNA recruitment into virions. Using crosslinking-immunoprecipitation (CLIP) sequencing, we uncover several dramatic changes in the RNA-binding properties of Gag that occur during virion genesis, coincident with membrane binding, multimerization, and proteolytic maturation. Prior to assembly, and after virion assembly and maturation, the nucleocapsid domain of Gag preferentially binds to ψ and Rev Response elements in the viral genome, and GU-rich mRNA sequences. However, during virion genesis, this specificity transiently changes in a manner that facilitates genome packaging; nucleocapsid binds to many sites on the HIV-1 genome and to mRNA sequences with a HIV-1-like, A-rich nucleotide composition. Additionally, we find that the matrix domain of Gag binds almost exclusively to specific tRNAs in the cytosol, and this association regulates Gag binding to cellular membranes.

INTRODUCTION

The HIV-1 Gag protein coordinates all major steps in virion assembly. In so doing, it changes subcellular localization and multimerization state and becomes proteolytically processed (Bell and Lever, 2013; Sundquist and Kräusslich, 2012). One function of Gag is to selectively package a dimeric, unspliced viral RNA genome selected from a pool of excess cellular RNAs and spliced viral mRNAs (Kuzembayeva et al., 2014; Lu et al., 2011b; Rein et al., 2011). Genome packaging requires binding of the nucleocapsid (NC) domain to viral genomic RNA (Aldovini and Young, 1990; Berkowitz et al., 1993; Gorelick et al., 1990). Selection of the viral genome is thought to be governed by a *cis*-acting packaging element, ψ (Ψ), within the 5' leader of the viral genome, composed of sequences in the unique 5' region (U5) and between the tRNA primer binding site (PBS) and the 5' portion of the Gag open reading frame (ORF) (Aldovini and Young, 1990; Clavel and Orenstein, 1990; Lever et al., 1989;

Luban and Goff, 1994). This element is highly structured (Clever et al., 1995; Harrison and Lever, 1992; McBride and Panganiban, 1996) and may exist in two conformations that favor translation versus dimerization and packaging (Lu et al., 2011a).

Knowledge of the viral RNA sequences that are directly bound by Gag is largely inferred from determinations of functional packaging signals in genetic studies, complemented by limited *in vitro* data. No assay has yet demonstrated a direct interaction between Gag and Ψ in a biologically relevant setting, i.e., in cells or virions. Additionally, some findings suggest that sequences outside Ψ might facilitate genome packaging. First, disruption of Ψ does not eliminate specific RNA encapsidation (Clever and Parslow, 1997; Laham-Karam and Bacharach, 2007; McBride and Panganiban, 1997). Second, sequences outside Ψ can increase HIV-1 vector titers or virion RNA levels (Berkowitz et al., 1995; Chamanian et al., 2013; Das et al., 1997; McBride et al., 1997; Richardson et al., 1993). Third, virions can package cellular RNAs (Muriaux et al., 2001; Rulli et al., 2007), lacking a Ψ sequence, particularly in the absence of viral RNA. The RNA properties that underlie these findings are unknown.

HIV-1 Gag molecules exist as monomers or low-order multimers in the cell cytosol and form higher-order multimers only after binding to the plasma membrane (Kutluay and Bieniasz, 2010). Imaging studies indicate that small numbers of Gag molecules recruit a single viral RNA dimer to the plasma membrane, nucleating the assembly of thousands of Gag molecules into an immature virion (Jouvenet et al., 2009). Thereafter, Gag proteolysis liberates NC and other Gag domains, triggering virion maturation. Whether these changes in Gag/NC configuration affect its RNA-binding properties is unknown, but there is clear potential for the RNA binding specificity of Gag to change during virion genesis.

Although NC is thought to be the primary Gag domain that binds to viral RNA, the matrix (MA) domain can also bind RNA *in vitro* (Alfadhli et al., 2009; Chukkapalli et al., 2010; Cimarelli and Luban, 1999; Levin et al., 2010; Ott et al., 2005; Ramalingam et al., 2011). N-terminal basic amino acids in MA that mediate membrane binding also drive *in vitro* RNA binding (Chukkapalli et al., 2010, 2013; Hill et al., 1996; Saad et al., 2006; Shkriabai et al., 2006; Zhou et al., 1994). Because RNA is better able to block MA binding to membranes that are devoid of phosphatidylinositol-(4,5)-bisphosphate PI(4,5)P₂ (Chukkapalli et al., 2010,

2013; Dick et al., 2013), RNA might help to target particle assembly to the plasma membrane. However, RNA binding by MA is not thought to be specific, and whether it actually occurs in cells is unknown.

To obtain a complete account of the RNA sequences bound by Gag during virion genesis, we employed crosslinking-immunoprecipitation-sequencing PAR-CLIP and HITS-CLIP techniques (Hafner et al., 2010; Licatalosi et al., 2008). We find that cytosolic Gag binds to three sequence elements within the 5' leader of the viral genome, which are brought together in a secondary structure that defines a minimal Ψ element. We also find that cytosolic Gag binds to additional discrete sites on the viral genome, including the Rev Response Element (RRE). Gag association with the plasma membrane and its assembly into immature virions trigger a profound change in RNA binding specificity that favors genome packaging. Subsequently, particle maturation largely reverses this change. Finally, we find that MA is a bona fide RNA binding domain that selects a subset of tRNAs in the cytosol, which regulate Gag-membrane binding. Overall, these studies provide a dynamic, quantitative, and high-resolution account of the global changes in Gag-RNA binding during HIV-1 virion genesis.

RESULTS

CLIP Assay for HIV-1 Gag-RNA Binding

We employed recently developed CLIP approaches (Hafner et al., 2010; Licatalosi et al., 2008) to identify RNA molecules bound by HIV-1 Gag protein during particle genesis. To facilitate the purification of Gag-RNA adducts, we generated HIV-1_{NL4-3} (subtype B) and HIV-1_{NDK} (subtype D) proviral clones carrying an inactivating mutation in the viral protease and three consecutive copies of a HA-tag within the stalk region of MA (MA-3xHA/PR⁻). The MA modification did not affect Gag expression or assembly and had only a small effect on the infectiousness of a PR⁺ virus (Figures S1A and S1B available online). Cells transfected with HIV-1_{NL4-3}(MA-3xHA/PR⁻) proviral plasmids were grown in the presence of ribonucleoside analogs (4SU or 6SG), which also had minimal effects on infectious virion yield (Figures S1C and S1D).

Cells and virions were UV-irradiated, lysed, and digested with ribonuclease A. Then, Gag-RNA adducts were immunopurified, end-labeled with γ -³²P-ATP, and visualized after SDS-PAGE and transfer to nitrocellulose membranes. Gag-RNA adduct formation was dependent on UV irradiation in cells and virions (Figure 1A). We primarily used 4SU-based CLIP thereafter, because it efficiently generated crosslinks, whereas 6SG and unmodified RNA was used for confirmatory purposes. Gag-crosslinked RNA oligonucleotides were purified, sequenced, and mapped to the HIV-1 and human genomes (see Extended Experimental Procedures). Reads derived from the terminal repeat (R) region of the HIV-1 genome ambiguously map to 5' and 3' ends but are displayed at 5' end of the viral genome, and cautiously interpreted, in our analyses.

In six independent Gag-CLIP libraries prepared from HIV-1_{NL4-3}(MA-3xHA/PR⁻) or HIV-1_{NDK}(MA-3xHA/PR⁻) expressing cells, 2.5%–7% of the total reads were HIV-1 derived, whereas ~60% were from host cell RNA (Figure 1B; Table S1). In compar-

ison, RNA sequencing (RNA-seq) libraries, which measure the abundance of cellular and viral RNAs, contained 0.3%–1.6% of reads derived from HIV-1, and ~75% were from cellular RNA (Figure 1B; Table S1). In immature virions, ~50% and 20% of CLIP reads were from viral and host RNAs, respectively (Figure 1B; Table S1), broadly similar to RNA abundance as determined by RNA-seq (Figure 1B). Thus, viral RNA sequences were somewhat selectively bound by Gag in cells but were enriched to a far greater extent in virions.

HIV-1 RNA Sequences Bound by Gag in Cells and Virions

We plotted the frequencies with which each nucleotide in the HIV-1_{NL4-3} genome was represented in reads from cell- and virion-derived HIV-1_{NL4-3}(MA-3xHA/PR⁻) Gag-CLIP libraries ("read density", Figure 1C). In cells, a major proportion of the Gag-linked reads were derived from discrete sites in the viral genome. As might be expected, the 5' leader was a frequent site of Gag binding. However, additional sites of frequent Gag binding included the RRE, sequences overlapping the Nef start codon, and untranslated sequences in U3 (Figure 1C). Several other sites on the viral genome were bound by Gag, but at lower frequencies. The distribution of Gag-crosslinked viral RNA reads was highly reproducible in HIV-1_{NL4-3}(MA-3xHA/PR⁻) Gag CLIP-seq libraries as indicated by the nearly perfect correlation between independent experiments (Figure 1D). However, when a divergent HIV-1_{NDK} strain was included in these analyses, the only prominent Gag binding sites that were present in both strains were the 5' leader and RRE (Figure 1C).

The frequency with which Gag was crosslinked to sites across the viral genome was starkly different in immature virions. Sites of frequent Gag binding on HIV-1_{NL4-3} RNA were far more numerous in immature virions than in cells. Moreover, specific Gag binding to the 5' leader and RRE was not evident in immature virions (Figure 1C). The pattern of Gag binding frequency across the viral genome was highly reproducible (Figure 1D) and was unlikely to be generated by methodological bias during CLIP-seq library generation, because it was largely unaffected by the choice of ribonucleoside analog (4SU versus 6SG; Figures S1E and S1F), ribonuclease (RNase A versus RNase T1; Figures S1G and S1H), or the immunoprecipitating antibody (anti-HA versus anti-NC; Figures S1I and S1J) used to generate the CLIP libraries. HIV-1_{NDK} Gag similarly bound to sites throughout the viral genome (Figure 1C). Despite overall similarity, there were some clear discrepancies in viral RNA sites that were frequently occupied by HIV-1_{NL4-3} and HIV-1_{NDK} Gag (Figures 1E and S1K), presumably due to differences in target RNA sequence, or subtle differences in the RNA binding specificities of the two Gag proteins.

The 3xHA tag did not affect the pattern of Gag binding at sites proximal or distal to the insertion site (Figure S1L). Moreover, autocorrelation analysis of Gag binding frequency revealed no peaks other than at a separation of $s = 0$ (Figure S1M). Thus, the frequency with which a given site in the viral genome was bound by Gag in immature virions appeared unaffected by its position relative to the Ψ sequence or other Gag binding sites. Rather, Gag binding appeared to be a function of local nucleotide sequence or structure.

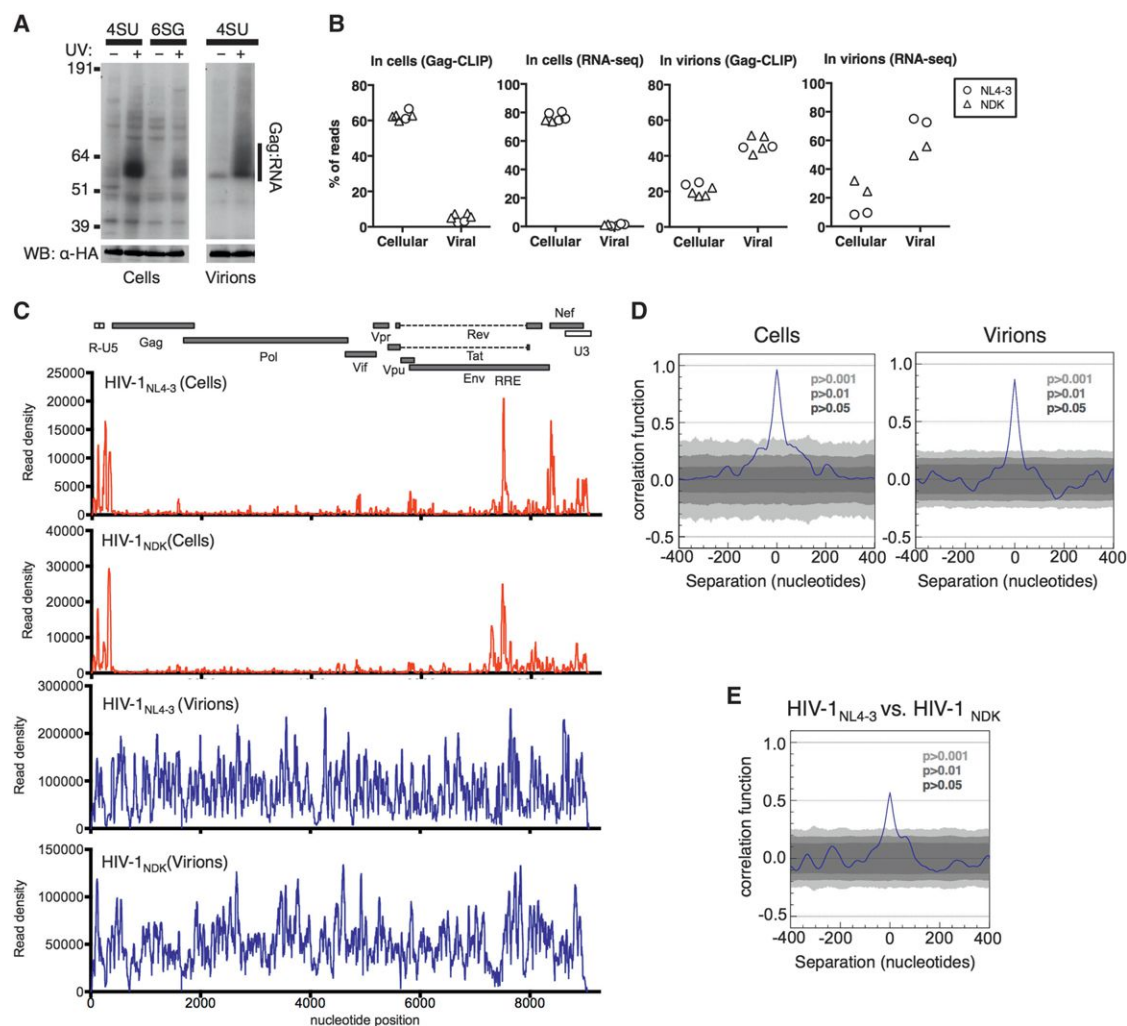


Figure 1. Viral RNA Binding by Gag in Cells and Virions

(A) Gag-RNA adducts immunoprecipitated from HIV-1_{NL4-3}(MA-3xHA/PR⁻)-expressing, 4SU- or 6SG-fed, cells or progeny virions, visualized by autoradiography (upper). Western blot analysis of the same membranes (lower).

(B) Proportion of reads that map to cellular and viral genomes obtained from Gag-CLIP and RNA-seq experiments performed on cell lysates and immature virions. Each point represents a separate library.

(C) Frequency distribution of nucleotide occurrence (read density) in reads mapping to viral genomes. A schematic diagram of HIV-1 genome features shown above is colinear.

(D) Correlation analysis of HIV-1_{NL4-3} Gag binding to viral RNA in cells and virions from independent CLIP experiments.

(E) Correlation analysis of HIV-1_{NL4-3} versus HIV-1_{NDK} Gag binding to viral RNA in virions.

See also Figure S1 and Table S1.

Analysis of Gag Binding to 5' Leader and RRE

In cells, the most prominent Gag binding sites on the viral genome coincided with the most prominently structured elements in the HIV-1 genome, namely, the 5' leader and the RRE (Figure 1C). However, Gag was not bound across the 5' leader and the RRE with uniform frequency, but selectively associated with small determinants within these structures. In the case of the 5' leader, Gag was most frequently bound to three distinct sequences, including one at the 5' end of U5, a second site between the PBS and the major splice donor, and a third site 3' to the major splice donor (Figure 2A). These sites are separated from each other in linear sequence by approximately 100 nucle-

otides. Strikingly, however, a nuclear-magnetic-resonance-based analysis of the structure of a Ψ sequence (Lu et al., 2011a) predicts that these three Gag-binding sequences would be in close proximity, and partly base paired with each other upon RNA folding into a structure that favors genome packaging (Figure 2B).

Frequent Gag binding to the RRE occurred in stem I, a site clearly distinct from the primary site of Rev-RRE interaction in stem-loop IIB (Malim et al., 1990) (Figures 2C and 2D). Importantly, the Gag-bound reads derived from both the 5' leader and RRE contained high rates of T-to-C substitution, identifying individual nucleotide bases that were in close proximity to Gag

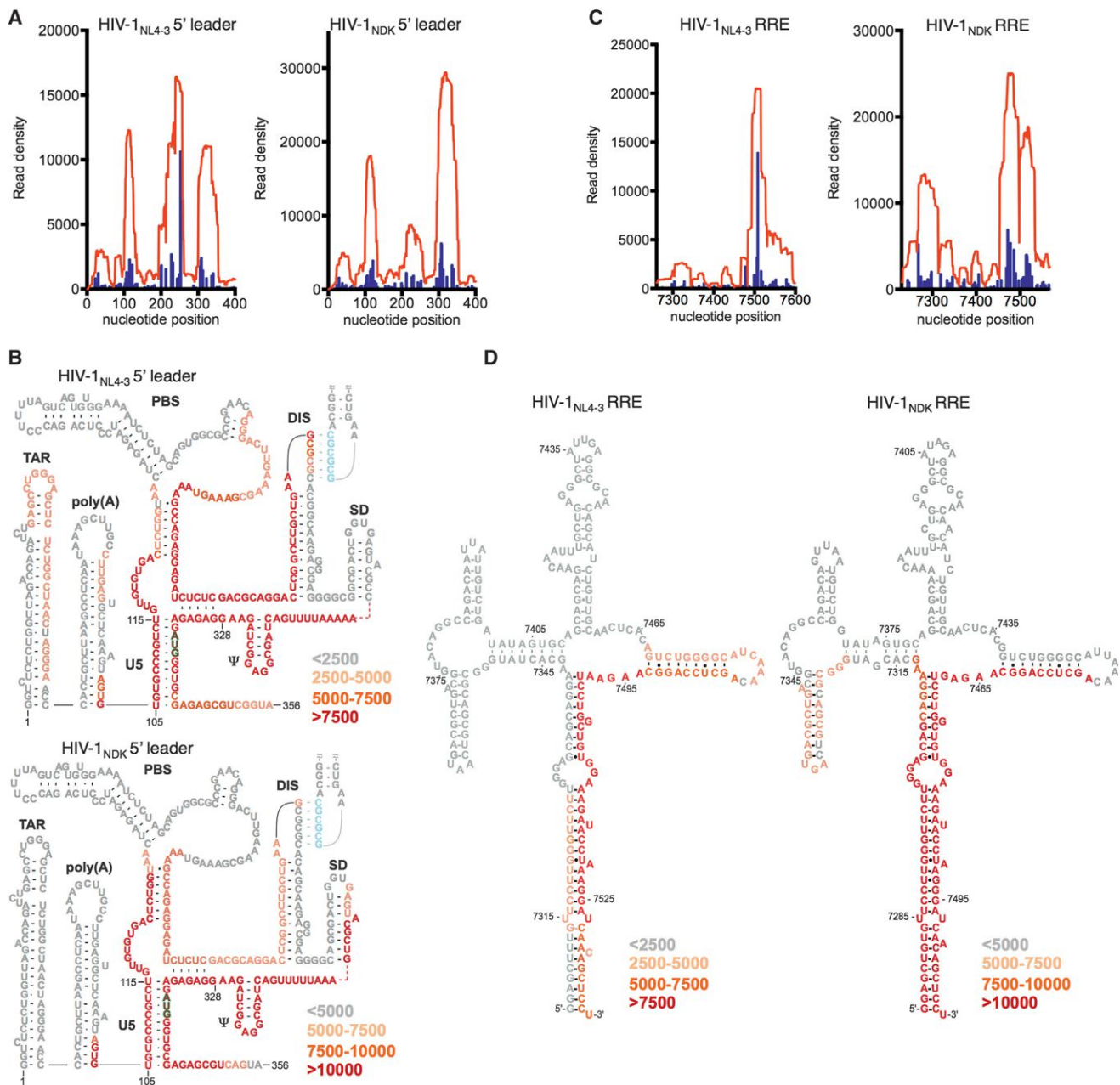


Figure 2. Details of HIV-1 Gag Binding to Ψ and RRE

(A and B) Read density distribution (red) and T-to-C mutations (blue) within the 5' leader (nucleotides 1–400, A) and RRE (B) in Gag-CLIP experiments with HIV-1_{NL4-3} and HIV-1_{NDK}.

(C and D) Depiction of read density distribution in the context of predicted secondary structures of the 5' leader (nucleotides 1–400, C) and RRE (D) in Gag-CLIP experiments with HIV-1_{NL4-3} and HIV-1_{NDK}. Color intensity represents read density indicated in the key.

See also Figure S2.

(Figures 2A and 2C). Additionally, CLIP-seq experiments performed using 4SU or 6SG yielded a similar footprint of Gag on viral RNA, with 5' leader and RRE binding prominently featured (Figures S2A and S2B). Thus, the apparently specific binding of Gag to these sequence elements was not an artifact of the particular crosslinking nucleotide used.

Changes in Gag RNA Binding to Viral RNA during HIV-1 Virion Assembly

Given the stark differences in the interaction of Gag with viral RNA in cells and immature virions, we attempted to determine what triggers this dramatic change. To this end, cells were fractionated after UV irradiation (Figure 3A), and Gag-RNA adducts

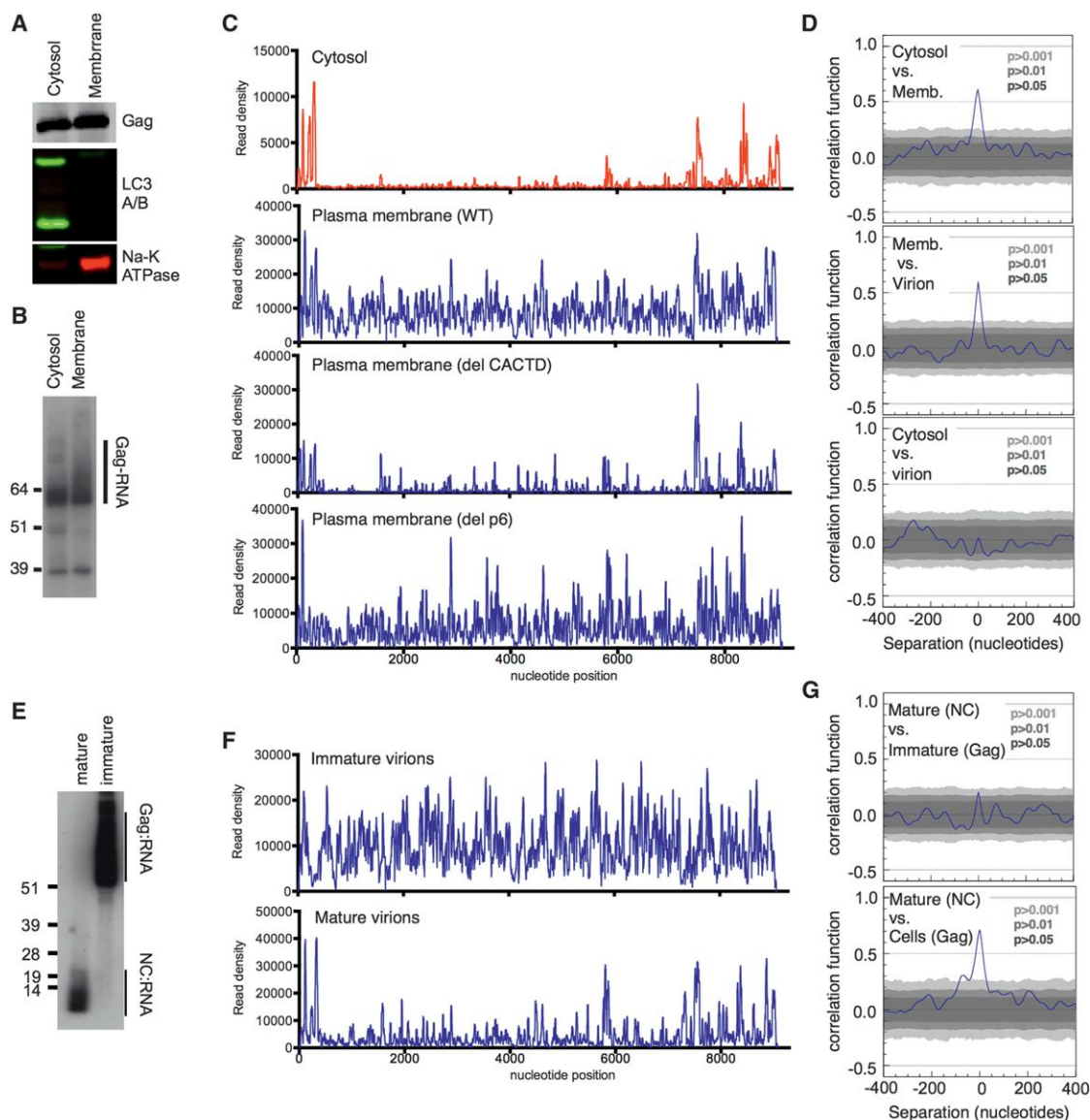


Figure 3. Changes in Gag Binding to Viral RNA during Virion Assembly and Maturation

(A) Western blot analysis of Gag, and markers of cytosol (LC3A/B) and plasma membrane (Na-K ATPase) in fractionated, 4SU-fed 293T cells transfected with a HIV-1_{NL4-3}(MA-3xHA/PR⁻) proviral plasmid.

(B) Autoradiogram of Gag-RNA complexes recovered from fractions in CLIP assays.

(C) Read density distribution on viral RNA from CLIP experiments in which WT and mutant Gag proteins were immunoprecipitated from cell fractions.

(D) Correlation analysis of CLIP data from cell fractions and immature virions.

(E) Autoradiogram of NC-RNA and Gag-RNA complexes recovered from mature and immature virions using an anti-NC antibody.

(F) Read density distribution on viral RNA from CLIP experiments in which Gag and NC proteins were immunoprecipitated from immature and mature virions using an anti-NC antibody.

(G) Correlation analysis of CLIP data from cells and mature and immature virions.

See also Figure S3.

were immunoprecipitated from cytosol, in which Gag is primarily monomeric, and from a membrane fraction where multimerized assembly intermediates form (Kutluay and Bieniasz, 2010) (Figure 3B). The RNA binding profile of cytosolic Gag was nearly indistinguishable from that described above for total cell lysates, with Ψ and RRE representing specific binding sites (Figures 3C

and 3D). Similarly, a myristoylation-defective Gag mutant (Gag-G2A), which does not bind efficiently to membranes, was also restricted to discrete binding sites on viral RNA (Figure S3A). Conversely, Gag that was immunoprecipitated from membrane fractions was bound to many sites on the viral genome, and, although binding to Ψ and RRE remained prominent, the overall

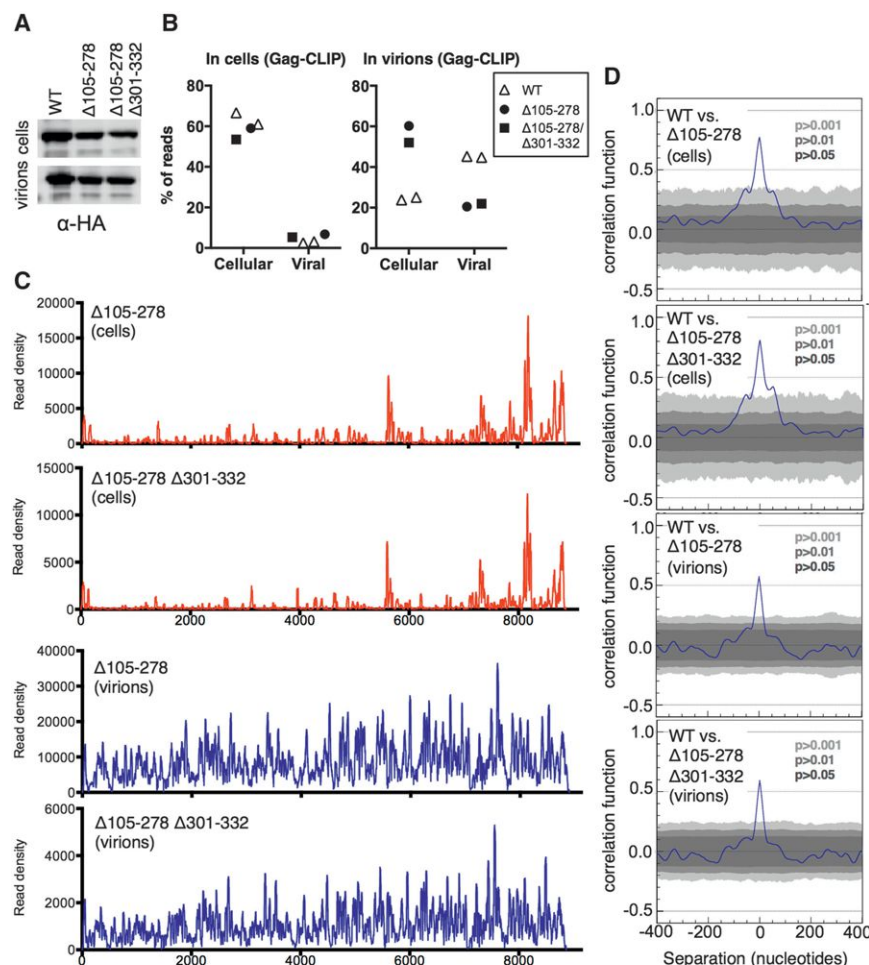


Figure 4. Gag Binding to Viral RNA in the Absence of Ψ

(A) Western blot analysis (anti-HA) of Gag protein expression from WT and Ψ -deleted ($\Delta 105-278$ and $\Delta 105-278/\Delta 301-332$) proviral plasmids.

(B) Proportion of reads from cell and virion CLIP experiments that map to cell and viral genomes, following transfection of WT or Ψ -deleted proviral plasmids. Each point represents a separate library.

(C) Read density distribution on viral RNA from CLIP experiments in which Gag proteins were immunoprecipitated from cell lysates and immature virions following transfection with Ψ -deleted proviral plasmids.

(D) Correlation analysis of CLIP data from cells and virions using WT and Ψ -deleted proviral plasmids. See also Figure S4.

RNA binding profile of cell-membrane-associated Gag otherwise resembled that of immature virion-associated Gag (Figures 3C, 3D, and S3B). A late-domain mutant Gag protein ($\Delta p6$) was similarly bound to sites throughout the viral genome (Figure 3C) indicating that the completion of budding was not required for this apparent shift in RNA binding specificity. Notably, a Gag CA Δ CTD mutant that is fully competent to localize to the plasma membrane but is significantly impaired in the formation of high-order multimers (Kutluay and Bieniasz, 2010) exhibited a binding profile that was more reminiscent of cytosolic Gag, although additional sites in the viral genome were bound with some prominence (Figure 3C). Together, these data indicate that high-order oligomerization of Gag at the plasma membrane drives a profound change in Gag's RNA-binding properties (Figure 3C, 3D, and S3B), dramatically increasing the extent to which Gag and viral RNA interact with each other.

Changes in Gag-Viral RNA Binding Triggered by Virion Maturation

Gag undergoes proteolytic cleavage during virion maturation, liberating NC, which is then thought to condense with the viral RNA inside a remodeled conical core (Sundquist and Kräusslich,

2012). To determine whether virion maturation affects Gag-viral RNA interactions, we did CLIP experiments using an antibody that specifically recognizes NC (Figures S3D and S3E) and could efficiently immunoprecipitate Gag-RNA and NC-RNA adducts from immature and mature virions, respectively (Figure 3E). Comparison of RNA sequences bound by Gag in immature virions and NC in mature virions revealed profound changes accompanying virion maturation (Figures 3F and 3G). Unlike intact Gag in immature virions, NC in mature virions preferentially occupied discrete sequences on the viral genome, that coincided in large part with the major sites of Gag-viral RNA interaction in the cytosol (i.e., Ψ and RRE) (Figures 3F and 3G). Many of the additional prominent NC-binding sites in mature virion RNA were also bound by intact Gag in the cytosol at lower frequencies (Figure S3F). Thus, the RNA-binding properties of NC in mature virions resemble that of unassembled Gag in the cytosol to a surprising extent, exhibiting statistically significant correlation (Figure 3G), possibly reflecting the monomeric state of cytosolic Gag and mature NC.

Gag Binding to Viral RNA and Viral RNA Packaging in the Absence of Ψ or RRE

To test the importance of the Gag/NC-binding sites in Ψ and RRE, we generated viruses carrying deletions of these sequences. First, we determined whether the Gag binding to additional sites on the viral genome in cells and in virions required initial binding to Ψ . To this end, we performed CLIP experiments using viral constructs carrying deletions of two of the three Gag-binding sequences ($\Delta 105-278$), or all three Gag binding sequences ($\Delta 105-278/\Delta 301-332$) within Ψ . These deletions left sequences surrounding the major 5' splice donor intact, and only modestly reduced Gag expression (Figure 4A). Surprisingly, in cells, neither the $\Delta 105-278$ nor the $\Delta 105-278/\Delta 301-332$

mutation affected the fraction of Gag-crosslinked reads derived from viral versus human genomes (Figure 4B). Moreover, in immature virions, these Ψ -deletions caused only a modest ~ 3 -fold decrease in the fraction of reads derived from the viral genome, with a corresponding increase in the fraction of reads from cellular RNAs (Figure 4B). The pattern of Gag binding to viral RNA in cells was not greatly affected by the Ψ deletions (Figures 4C and 4D), although there was an apparent reduction in the frequency with which RRE was bound. Similarly, the overall pattern of Gag binding to viral RNA in immature virions was also relatively unperturbed by Ψ deletions (Figures 4C, 4D, and S4A). However, there was a tendency for read densities to increase toward the 3' end of the viral genome, perhaps reflecting the incorporation of spliced viral RNAs into virions generated by Ψ -deleted genomes (Clever and Parslow, 1997; Houzet et al., 2007; Russell et al., 2003). Importantly, these results indicate that multiple sites on the viral genome bind to Gag independently of Ψ and facilitate HIV-1 genome packaging.

To analyze the effects of Gag-RRE interaction on HIV-1 infectivity, we generated viruses in which Rev was deleted and the nuclear export of unspliced viral RNAs was mediated by the Mason-Pfizer monkey virus constitutive transport element (CTE) to circumvent effects of RRE on RNA nuclear export (Figure S4B). In this setting, the presence or absence of RRE had no discernible effect on Gag protein levels or infectious virus yield (Figure S4C). Although it is formally possible that the MPMV CTE might recapitulate the Gag-binding properties of RRE, this analysis suggests that Gag RRE interactions do not regulate Gag expression and are not required for RNA packaging.

Changes in the Sequence Specificity of Gag RNA Binding during Virion Genesis

The aforementioned results suggest a degree of redundancy, and contributions from multiple domains of the viral genome to RNA packaging. In an attempt to explain the factors driving the selectivity with which Gag bound to and packaged viral RNA, we determined the identities of cellular mRNA sequences that were most frequently bound to Gag in cells and in virions. Reads that aligned to the human genome were clustered using PARalyzer (Corcoran et al., 2011), which defines a cluster, or binding site, based on the occurrence of a minimum number overlapping reads proximal to a T-to-C substituted crosslinking site in 4SU-based CLIP assays (Table S2). Note that a single cluster is counted once for the analysis shown in Figure 5A, irrespective of the number of reads associated with it. In cells, $>90\%$ of clusters (Gag binding sites) were within genes, and $\sim 80\%$ of these were derived from mRNAs (Figure 5A) with an overlap of $>95\%$ between HIV-1_{NL4-3} and HIV-1_{NDK} Gag-bound mRNA clusters (Figure S5A). A similar analysis of immature virions generated a collection of clusters that were also mostly derived from mRNAs (Figure 5A) and overlapped by $>70\%$ in HIV-1_{NL4-3} and HIV-1_{NDK} libraries. Notably, greater discrepancies were observed when the Gag-binding clusters in cells versus virions were compared (Figure S5A), indicating that many mRNA sequences that are preferred Gag binding sites in cells are not preferred binding sites in virions and vice versa.

We counted the number of reads associated with each cluster and determined the 100 sites in cellular mRNAs that were most

frequently bound by Gag in cells and immature virions. The nucleotide composition of these Gag binding sites revealed a striking change in the RNA binding preference of Gag during virion genesis (Figure 5B). In cells, preferred Gag binding sites had a strong tendency to be G-rich (mean G-content of $\sim 40\%$). In contrast, in immature virions, preferred Gag binding sites displayed a tendency to be A-rich. Remarkably, the nucleotide content of the preferred Gag binding sites in cellular mRNAs associated with immature virions was strikingly similar to the nucleotide composition of the HIV-1 genome.

We used cERMIT (Georgiev et al., 2010), to search for sequence motifs that were enriched in Gag-bound cellular mRNA clusters recovered at various stages of virion genesis. G/U-rich sequence motifs were the most often present in host mRNA sequences bound to Gag in the cytoplasm (Figure 5C). At the plasma membrane, and particularly in immature virions, there was a clear change in binding specificity. Here, Gag bound more frequently to A-rich and A/G-rich cellular mRNA motifs (Figure 5C). Inspection of the sequences of the 12 clusters most frequently bound by Gag in the viral genome, in immature virions, revealed multiple instances of similar A/G-rich motifs scattered in the viral genome (Figure 5D). In mature virions, RNA binding specificity reverted, and the sequence motifs favored by NC were again G/U-rich. These findings reinforce the notion that Gag binds preferentially to RNAs that have an A-rich nucleotide composition, and particularly to A-rich motifs that are present in the viral genome, transiently during immature virion assembly.

tRNAs Are the Most Frequently Gag-Bound RNA in Cells, but Not in Virions

Although the aforementioned analysis focused on Gag binding to mRNA sequences, when the number of individual Gag-bound reads associated with each cluster or binding site was counted, it was evident that mRNAs were responsible for only a fraction ($\sim 12\%$) of the cellular RNA reads crosslinked to Gag. In fact, tRNAs were the dominant RNA species bound by Gag in cells ($\sim 60\%$ – 70% of reads), with 7SL RNA, a known component of HIV-1 virions (Onafuwa-Nuga et al., 2006) constituting the bulk of the remaining reads (Figure 6A). Strikingly, the fraction tRNA-derived reads was greatly reduced in CLIP-seq libraries prepared from virions as compared to cells. This was not the case for other RNA classes, suggesting that tRNAs were specifically excluded from Gag-bound RNAs in virions (Figure 6A).

MA Binds to Particular tRNAs

Although the aforementioned Gag-RNA binding events were assumed to involve the NC domain of Gag, there are numerous reports that MA can interact with viral or cellular RNAs (reviewed in Alfadhli and Barklis, 2014). However, these interactions, generally assayed *in vitro*, are of uncertain significance. To test whether MA binds to RNA in cells, we modified the aforementioned HIV-1_{NL4-3}(MA-3xHA/PR⁻) clones to include a FXa protease cleavage site immediately N- or C-terminal to the 3xHA-tag (Figures 6B and S6A). After UV crosslinking, cell lysates could be treated with FXa protease to generate two Gag fragments (MA and CA-NC-p6) only one of which would be tagged and immunoprecipitated (Figures 6B and 6C).

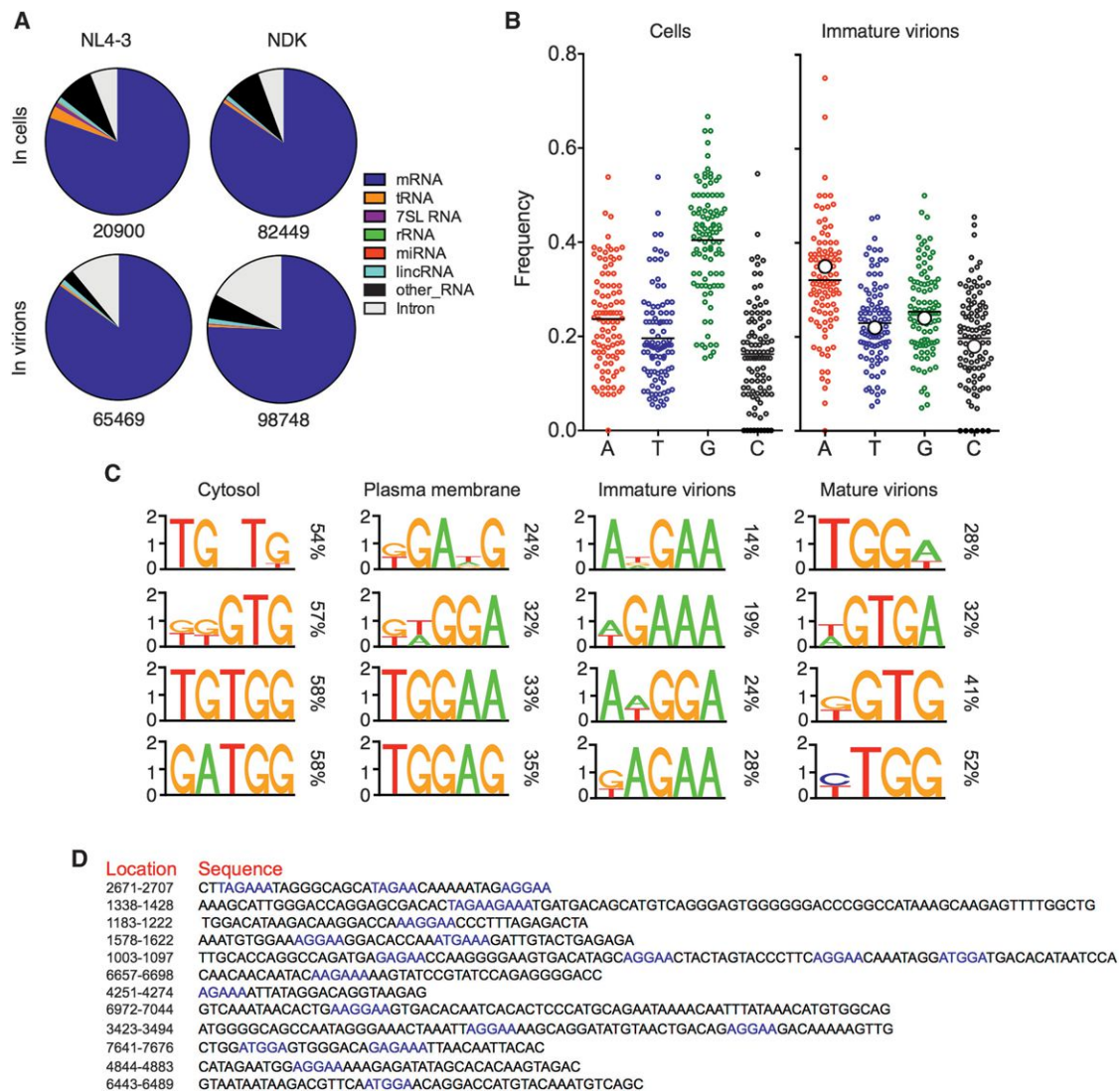


Figure 5. Changes in Gag-RNA Binding Specificity during Virion Genesis

(A) Classification of PARalyzer-generated read clusters that map to cellular RNAs; total number of clusters is indicated.

(B) Nucleotide composition of 100 PARalyzer-generated clusters with largest numbers of reads, derived from Gag-bound host mRNAs in cells and immature virions. Larger white circle symbols indicate nucleotide composition of the HIV-1 genome.

(C) Sequence motifs identified by cERMIT as most frequently occurring in Gag bound, host cell mRNA clusters at various steps of virion genesis. Cumulative percentages of clusters containing the motifs are indicated.

(D) Sequences of the 12 viral genome-derived clusters that are most frequently Gag-bound in immature virions. Motifs matching those identified as preferred binding sites in host mRNAs are highlighted in blue.

See also Figure S5 and Table S2.

Like full-length Gag, the immunoprecipitated CA-NC-p6 contained crosslinked RNA species that caused adducts to migrate ~0–10 kDa above the expected molecular weight of the protein. Notably, the isolated MA domain also generated prominent crosslinked species, indicating RNA binding. Unusually, these MA-RNA adducts migrated at a discrete molecular weight of ~35 kDa, suggesting that MA bound to a unique RNA species (Figure 6D).

Although the CA-NC-p6-RNA adducts contained viral RNA sequences with a very similar distribution to those recovered from

full-length Gag, MA-RNA adducts were nearly devoid of viral RNA-derived sequence (Figure 6E). In fact, the RNA molecules crosslinked to MA were almost exclusively host tRNAs (Figure 6F). When RNase treatment was omitted from the CLIP procedure prior to immunoprecipitation, MA-RNA adducts migrated at a larger, but still discrete, molecular weight (Figure 6D) matching the size of one tRNA molecule (~20 kDa) crosslinked to MA, suggesting that MA binds to full-length tRNAs.

Although tRNAs constituted the majority of cellular RNA molecules bound by intact Gag in cells, deletion of MA caused a major

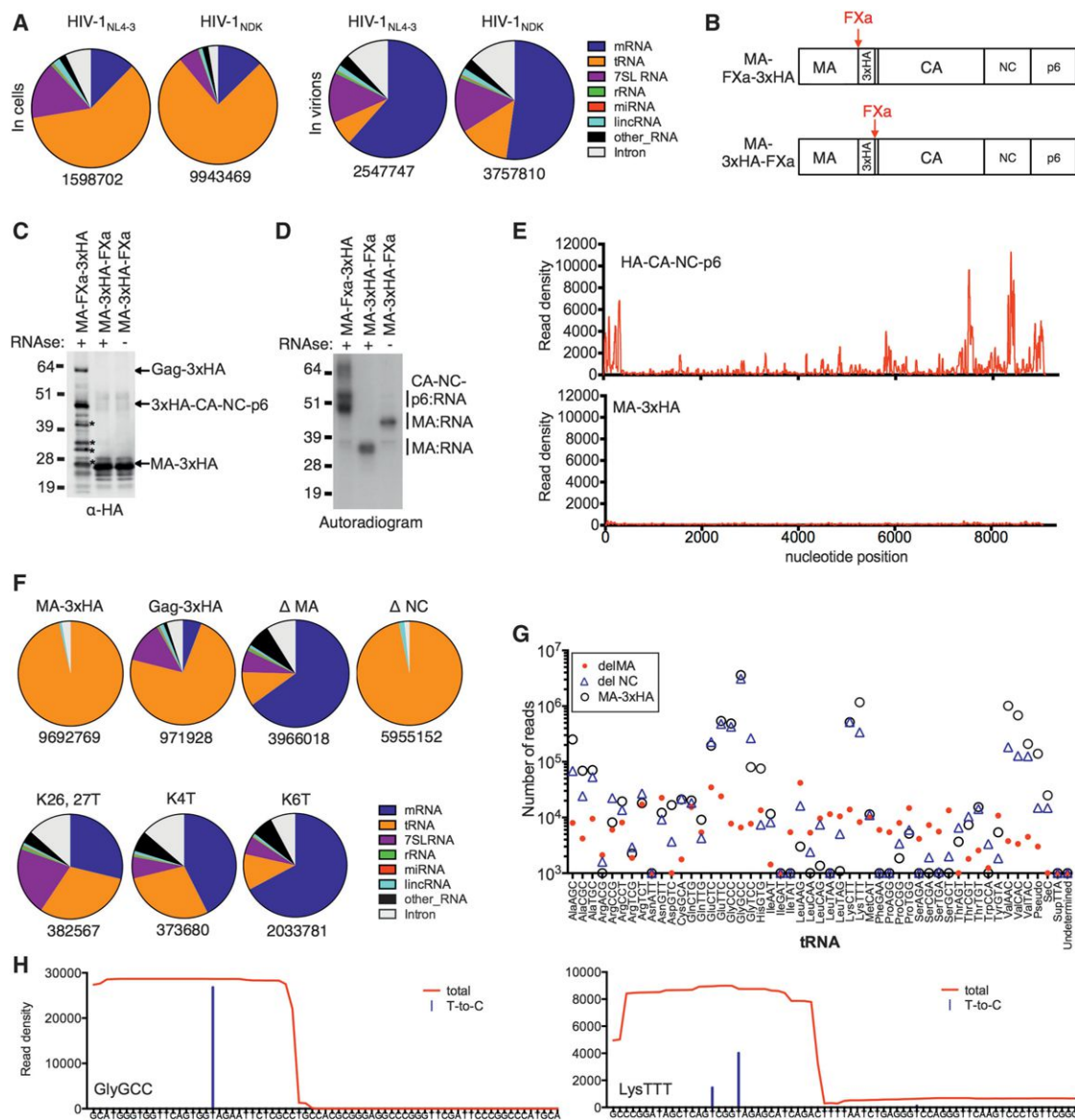


Figure 6. HIV-1 MA Binds to Specific tRNAs, but Not Viral RNA

(A) Classification of individual reads from cell and immature virion Gag-CLIP experiments that map to genes in the host cell genome.

(B) Schematic representation of FXa-cleavable Gag proteins.

(C and D) Western blot analysis (anti-HA, C) and autoradiography (D) of immunoprecipitated (anti-HA) proteins from HIV-1_{NL4-3} MA-FXa-3xHA/PR⁻ and MA-3xHA-FXa/PR⁻-transfected cells that were subjected to the modified CLIP procedure. Cell lysates were treated with Factor Xa protease, with or without RNase A prior to immunoprecipitation. *Breakdown products of the 3xHA-CA-NC-p6 protein that retain the HA tag but are not crosslinked to RNA.

(E) Read density distribution on viral RNA following CLIP analysis of FXa-liberated 3xHA-CA-NC-p6 and MA-3xHA proteins.

(F) Classification of individual reads that map to cellular genes from cell associated Gag-CLIP experiments employing the FXa-liberated MA-3xHA protein or WT and mutant, uncleaved Gag proteins. Total number of reads is indicated below.

(G) Numbers of reads crosslinked to FXa-liberated MA-3xHA or uncleaved Gag proteins that lack NC (ΔNC) and MA (ΔMA) that map to each tRNA gene.

(H) Two examples of read density distributions (red) and T-to-C substitutions (blue) at crosslinking sites on tRNAs (Gly[GCC] and Lys[TTT]) following CLIP analysis of FXa-liberated MA-3xHA.

See also Figure S6.

reduction in the fraction of reads that were tRNA derived but did not inhibit binding to mRNAs or 7SL RNA (Figures 6F and S6B). Conversely, a mutant Gag protein lacking the NC domain did not bind mRNA or 7SL, but bound nearly exclusively to tRNAs (Fig-

ures 6F and S6B). Progressive and specific loss of tRNA binding occurred as more lysines in the N-terminal MA basic domain were substituted, and mutation of six lysines (MAK6T) reduced Gag-tRNA binding to the same degree as deletion of MA (Figures

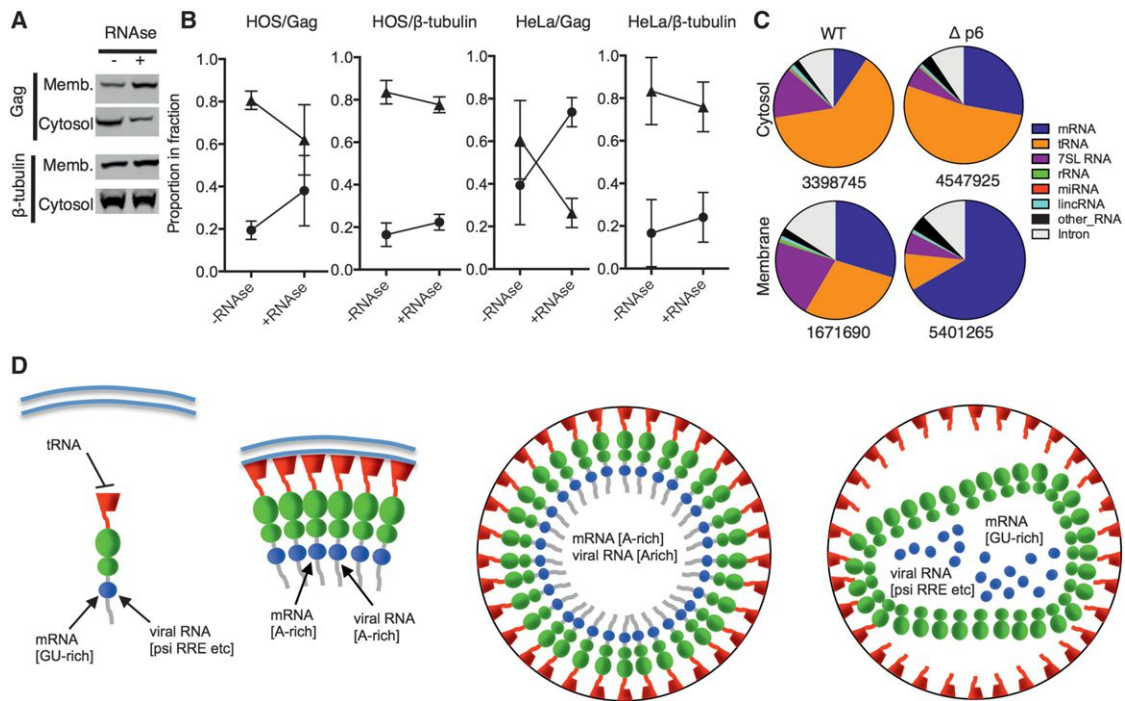


Figure 7. Reciprocal Relationship between HIV-1 Gag Binding to RNA and Membrane during Virion Genesis

(A) Western blot analysis (anti-CA and anti-β-tubulin) of HOS cell lysates, fractionated by membrane flotation after treatment with RNase. (B) Quantitation of Gag and β-tubulin in the cytosol and membrane fractions from five independent experiments each with HOS and HeLa cells. (C) Classification of individual reads that map to genes in the host cell genome from Gag-CLIP experiments done using fractionated cells and WT or mutant Gag proteins as described in Figure 3. Total number of reads is indicated below. (D) Model summarizing changing interactions between Gag and RNA during virion genesis.

6F and S6B). As expected, MAK4T and MAK6T mutations caused decreases in the levels of released particles (Figure S6B), consistent with a plasma membrane targeting defect in these mutants.

MA binding to tRNAs was highly selective. Indeed, GluCTC, GluTTC, GlyGCC, GlyCCC, LysCTT, LysTTT, ValAAC, and ValCAC tRNAs were bound up to 100-fold more frequently than the majority of tRNAs (Figure 6G). No such enrichment occurred in CLIP experiments done with GagΔMA. Although intact tRNA molecules were bound by MA (Figure 6D), binding apparently involved the 5' half of the tRNA molecules, and especially the dihydrouridine loop, as indicated by the very high rates of T-to-C conversions at this site (Figure 6H). Binding was not induced by 4SU incorporation into tRNAs, because the same tRNAs were selectively MA-bound in CLIP experiments where crosslinking was induced by UV irradiation at 254 nm in the absence of modified nucleotides (Figure S6C).

Regulation of Gag Membrane Binding by tRNA

The finding that the same lysine residues that mediate plasma membrane binding also mediate tRNA binding, suggested that tRNAs might regulate Gag localization. Indeed, previous work has indicated that exogenous RNA can inhibit Gag binding to liposomes *in vitro* (Chukkapalli et al., 2010, 2013; Dick et al., 2013). We took a simple cell-based approach (without addition of exogenous RNA or liposomes) and tested whether RNase

treatment of cell lysates increased Gag binding to endogenous cellular membranes. Lysates of HeLa and HOS cells stably expressing Gag-CFP were treated with ribonucleases and subjected to membrane flotation analysis. Strikingly, RNase treatment caused a significant redistribution of Gag from cytosol to membrane fractions (Figures 7A and 7B), whereas a control protein, β-tubulin, was largely unaffected. Consistent with a model in which tRNAs compete with membranes for Gag binding, tRNAs comprised a significantly smaller fraction of Gag-bound RNAs at the plasma membrane than in the cytoplasm (Figure 7C). This difference was more pronounced when particle budding was blocked at the plasma membrane by deletion of the late budding domain of Gag (Figure 7C). Thus, these results indicate that MA can bind to tRNA or cell membranes, but not both simultaneously, and strongly suggest that tRNAs regulate Gag localization by binding to basic amino acids in MA.

DISCUSSION

Two central conclusions of these studies are (1) the HIV-1 Gag protein has two RNA binding domains (NC and MA) with very different specificities and (2) dramatic changes in RNA binding regulate Gag localization and genome packaging during virion genesis (Figure 7D).

Prior to virion assembly, Gag exists as a diffuse pool of monomers or low-order multimers in the cytoplasm with its NC domain

bound primarily to mRNA, with some binding to 7SL and tRNA (Figure 7D). A fraction of NC-RNA interactions are with particular sites on the HIV-1 genome, including the Ψ sequence. Our data reveal the specific RNA sequences within Ψ that are in proximity to Gag, in a physiological setting. Satisfyingly, the three noncontiguous RNA elements that are most frequently crosslinked to Gag (nucleotides ~100–126, 195–260, and 300–350) coincide nearly precisely with a minimal Ψ element, which adopts a secondary structure that putatively favors genome packaging (Lu et al., 2011a).

Surprisingly, cytoplasmic Gag bound to additional discrete elements on the viral RNA, including RRE. Previously, *env* sequences, including RRE, were shown to facilitate packaging (Kaye et al., 1995; Richardson et al., 1993), but a discrete packaging sequence within *env* could not be mapped. Other reports suggest that Rev enhances packaging (Brandt et al., 2007), although effects of Rev/RRE on viral RNA nuclear export are potential confounders in packaging experiments. Despite strong evidence for Gag-RRE binding, the RRE did not affect infectious virion yield when its nuclear export function was replaced, suggesting that Gag-RRE binding is not required for genome packaging. Although it is possible that Gag/NC-RRE interaction plays a redundant role in packaging, other plausible functions include (1) shielding double-stranded stem I RNA from cytoplasmic sensors, (2) coupling RNA-export with packaging, (3) displacement of Rev from the RRE for recycling, (4) regulation of Env translation, and (5) enhancement of reverse transcription via NC's unwinding/chaperone function (Levin et al., 2010). Further work will be required to elucidate the functional role, if any, of Gag/NC-RRE binding.

In the cytosol, Gag favored binding to the viral genome over cellular mRNAs by a few fold. This level of discrimination is insufficient to account for the selectivity with which viral genomes are packaged. Our data indicate that a dramatic change in Gag-RNA binding specificity, coincident with CA-CTD-dependent high-order multimerization at the plasma membrane, contributes to selective packaging (Figure 7D). GU-rich sequences in cellular mRNA were targeted by Gag in the cytosol, consistent with previous *in vitro* and structural studies indicating that the isolated, monomeric NC domain favors binding to such sequences (Berglund et al., 1997; De Guzman et al., 1998; Fisher et al., 1998). However, during assembly, Gag molecules become tightly packed in hexameric lattices (Briggs et al., 2009; Wright et al., 2007). By constraining thousands of NC domains into a pseudo-2D curved array, local NC concentration is dramatically elevated. Potentially, features of NC that govern RNA binding specificity (zinc knuckles and basic amino acids) might be differently accessible in an assembled Gag lattice. Under these conditions, we found that A-rich mRNA sequences were preferentially bound by Gag (Figure 7D). Remarkably, the nucleotide composition of mRNA sequences bound by assembled Gag reflects an unusual, heretofore unexplained, property of the HIV-1 genome. Thus, our findings suggest that a need to selectively package viral RNA caused HIV-1 to evolve an unusually A-rich genome. Conversely, an A-rich genome may have evolved for other reasons and then drove Gag to gain a unique oligomerization-driven specificity for A-rich RNA. Notably, Ψ -deletion caused only a 3-fold reduction in the fraction of virion-associated,

Gag-bound, RNA sequences that were viral RNA derived. Moreover, the pattern of Gag binding to many sites in the viral genome was not solely a secondary effect of physical proximity to Gag- Ψ interactions. Rather, it appears that HIV-1 genome packaging is a two-step process, involving interactions between (1) Ψ and monomeric Gag and (2) A-rich viral RNA and multimeric Gag. This scenario should selectively drive particle assembly on viral RNAs, and we speculate that the biases in nucleotide composition exhibited by HIV-1 and other retroviruses serves as a proofreading-like mechanism to enhance the fidelity of genome packaging following initial Gag- Ψ interaction.

An unexpected finding was that proteolytic cleavage of Gag caused NC to revert to a preference for GU-rich mRNAs and discrete viral RNA sequences. This result reinforces the notion that Gag/NC RNA binding specificity is multimerization dependent. By liberating the majority of viral RNA from NC, while maintaining interaction with structured elements (to enable NC's chaperone activity), maturation-dependent changes in Gag/NC-RNA binding could facilitate reverse transcription (Levin et al., 2010).

Another surprising finding was that MA binding to specific tRNAs constitutes the most frequent binding events between Gag and RNA in cells. MA-tRNA binding was independent of NC and the PBS and is thus unlikely to involve the tRNA primer annealed to viral genome. Rather, we found that MA-tRNA interaction could regulate the binding of Gag to cellular membranes. MA specifies the location of virion assembly, and it was previously shown that RNA can block *in vitro* MA binding specifically to liposomes that lack acidic phospholipids (Chukkapalli et al., 2010, 2013). Thus, occlusion of MA basic residues by specific tRNAs might inhibit nonproductive assembly at most intracellular membranes and facilitate targeting to the plasma membrane where resident lipids have a high affinity for MA. Alternatively, MA-tRNA binding might provide a mechanism by which virion assembly is temporally regulated.

MA-tRNA interactions could serve additional purposes. Some degree of RNA binding may be an inevitable consequence of encoding a highly basic domain. Thus, specific MA binding in a 1:1 complex to small RNAs might be a mechanism to avoid the aggregation of a protein that has both two distinct RNA binding domains and a tendency to multimerize. MA-tRNA binding might also prevent nonproductive interaction of a viral genome with a Gag monomer whose NC domain has engaged viral RNA. Conceivably, MA may facilitate the selection of the RT primer, as tRNA^{Ala}3 is among the tRNAs bound by MA, but several other tRNAs are also bound by MA more frequently. Finally, MA-tRNA interaction could regulate viral and/or host translation. The A-richness of the HIV-1 genome results in suboptimal codon usage (Grantham and Perrin, 1986; Kypr and Mrázek, 1987; Sharp, 1986) and an elevated number of Ile, Lys, Glu, and Val codons in the Gag and Pol ORFs (Berkhout and van Hemert, 1994). Notably, Lys, Glu, and Val tRNAs are among those specifically bound by MA, providing a potential opportunity for Gag to regulate its own translation as it accumulates to high levels and sequesters tRNAs (perhaps facilitating packaging as a consequence). Similarly, MA could inhibit translation of host mRNAs whose products may be deleterious for viral replication. Indeed, interaction of MA with host elongation factors via a tRNA bridge

has been reported to inhibit translation in vitro (Cimarelli and Luban, 1999).

Overall, our global survey reveals surprising ways in which the interaction between Gag, viral, and host RNAs can change and modulate the process of virion genesis and genome packaging.

EXPERIMENTAL PROCEDURES

Proviral Plasmids and Cells

HIV-1_{NL4-3}-derived proviral plasmids containing a 3xHA tag in the stalk region of matrix (HIV-1_{NL4-3}MA-3xHA) were constructed using overlap extension PCR. Various derivatives of this construct encoding a catalytically inactive viral protease (PR⁻), a Factor Xa cleavage site on either side of the HA tag (MA-FXa-3xHA and MA-3xHA-FXa), deletions of nucleocapsid domain (Δ NC), the globular head of MA (Δ MA), the CA CTD (CA Δ CTD), or the Ψ signal (Δ 105-278 and Δ 105-278/ Δ 301-332) were constructed using PCR-based deletion mutagenesis. Constructs carrying mutations at binding sites for Tsg101 and ALIX proteins in the p6 domain of Gag (Δ p6), the Gag myristoylation signal (G2A), or at lysine residues in MA (MAK26,27T MAK4T and MAK6T) were constructed using PCR overlap extension-based mutagenesis. Proviral plasmids with deletions or mutations in Env, Rev, and RRE and encoding a Mason-Pfizer monkey virus constitutive transport element were similarly constructed. Details of the construction are described in Extended Experimental Procedures.

PAR-CLIP, HITS-CLIP, and RNA-Seq

For CLIP experiments, HEK293T cells were grown in 10 cm dishes and transfected with proviral plasmids using polyethylenimine (PolySciences). Virions were harvested from filtered supernatant by ultracentrifugation through sucrose and UV irradiated, whereas cells were irradiated while adhered to culture dishes. Prior to UV crosslinking, a fraction of cells and virions were collected for RNA-seq analysis. After crosslinking, the CLIP procedure was performed on unfractionated lysates (after removal of nuclei) or on membrane and cytoplasmic fractions.

For the CLIP procedure, cell and virion lysates were treated with RNaseA or RNaseT1 and then incubated with Protein G-conjugated Dynabeads coated with anti-HA or anti-NC antibodies. After immunoprecipitation of RNA-protein adducts, beads were washed and treated sequentially with calf intestinal alkaline phosphatase and then polynucleotide kinase and γ -³²P-ATP. RNA-protein adducts were eluted from the beads, separated by SDS-PAGE, blotted onto nitrocellulose, and digested with proteinase K. RNA oligonucleotides were then ligated to adapters, amplified by PCR and sequences determined using a Illumina HiSeq 2000 platform.

CLIP-seq experiments were performed two to six times on cells and virions. Further details of the method are in the Extended Experimental Procedures.

Bioinformatic Analyses

The FASTX toolkit was used to process the reads from CLIP and RNA-seq libraries before mapping. Reads were separated based on their 5' barcode sequences and collapsed to generate a set of unique sequences. Unique CLIP-seq and RNA-seq reads were mapped to the human (hg19) and HIV-1 genomes using the Bowtie. SAMtools, in house scripts, and GraphPad Prism were used to calculate and display read densities associated with viral and cellular RNAs and to calculate correlations. Clusters were generated using PARalyzer and used as an input for the CERMIT motif finding tool. Further details are given in the Extended Experimental Procedures.

Membrane Flotation Assays

The membrane flotation assays were performed using HeLa and HOS cells stably expressing cyan fluorescent protein-tagged Gag proteins as described (Kutluay and Bieniasz, 2010) with modifications outlined in the Extended Experimental Procedures.

ACCESSION NUMBERS

CLIP-seq data were deposited to the GEO database with accession number GSE61508.

SUPPLEMENTAL INFORMATION

Supplemental Information includes Extended Experimental Procedures, six figures, and two tables and can be found with this article online at <http://dx.doi.org/10.1016/j.cell.2014.09.057>.

AUTHOR CONTRIBUTIONS

S.B.K. and P.D.B. conceived the study, designed the experiments, analyzed the data, and wrote the paper. S.B.K., T.Z., C.P., and D.J. executed the experiments. S.B.K., D.B.M., and M.E. analyzed the sequencing data. P.D.B. supervised the work.

ACKNOWLEDGMENTS

We thank Markus Hafner and members of the Darnell and Tuschl labs for advice and helpful discussions. We thank Michael Summers for the diagram of the Ψ and RRE sequences in Figure 2. S.B.K. was supported in part by an AmFAR Mathilde Krim Postdoctoral Fellowship. This work was supported by NIH grants R01AI01111 and P50GM103297.

Received: August 3, 2014

Revised: September 17, 2014

Accepted: September 24, 2014

Published: November 6, 2014

REFERENCES

- Aldovini, A., and Young, R.A. (1990). Mutations of RNA and protein sequences involved in human immunodeficiency virus type 1 packaging result in production of noninfectious virus. *J. Virol.* 64, 1920–1926.
- Alfadhli, A., and Barklis, E. (2014). The roles of lipids and nucleic acids in HIV-1 assembly. *Front. Microbiol.* 5, 253.
- Alfadhli, A., Still, A., and Barklis, E. (2009). Analysis of human immunodeficiency virus type 1 matrix binding to membranes and nucleic acids. *J. Virol.* 83, 12196–12203.
- Bell, N.M., and Lever, A.M. (2013). HIV Gag polyprotein: processing and early viral particle assembly. *Trends Microbiol.* 21, 136–144.
- Berglund, J.A., Charpentier, B., and Rosbash, M. (1997). A high affinity binding site for the HIV-1 nucleocapsid protein. *Nucleic Acids Res.* 25, 1042–1049.
- Berkhout, B., and van Hemert, F.J. (1994). The unusual nucleotide content of the HIV RNA genome results in a biased amino acid composition of HIV proteins. *Nucleic Acids Res.* 22, 1705–1711.
- Berkowitz, R.D., Luban, J., and Goff, S.P. (1993). Specific binding of human immunodeficiency virus type 1 gag polyprotein and nucleocapsid protein to viral RNAs detected by RNA mobility shift assays. *J. Virol.* 67, 7190–7200.
- Berkowitz, R.D., Hammarskjöld, M.L., Helga-Maria, C., Rekosh, D., and Goff, S.P. (1995). 5' regions of HIV-1 RNAs are not sufficient for encapsidation: implications for the HIV-1 packaging signal. *Virology* 212, 718–723.
- Brandt, S., Blissenbach, M., Grewe, B., Konietzny, R., Grunwald, T., and Uberla, K. (2007). Rev proteins of human and simian immunodeficiency virus enhance RNA encapsidation. *PLoS Pathog.* 3, e54.
- Briggs, J.A., Riches, J.D., Glass, B., Bartonova, V., Zanetti, G., and Kräusslich, H.G. (2009). Structure and assembly of immature HIV. *Proc. Natl. Acad. Sci. USA* 106, 11090–11095.
- Chamanian, M., Purzycka, K.J., Wille, P.T., Ha, J.S., McDonald, D., Gao, Y., Le Grice, S.F., and Arts, E.J. (2013). A cis-acting element in retroviral genomic RNA links Gag-Pol ribosomal frameshifting to selective viral RNA encapsidation. *Cell Host Microbe* 13, 181–192.
- Chukkappalli, V., Oh, S.J., and Ono, A. (2010). Opposing mechanisms involving RNA and lipids regulate HIV-1 Gag membrane binding through the highly basic region of the matrix domain. *Proc. Natl. Acad. Sci. USA* 107, 1600–1605.

- Chukkappalli, V., Inlora, J., Todd, G.C., and Ono, A. (2013). Evidence in support of RNA-mediated inhibition of phosphatidylserine-dependent HIV-1 Gag membrane binding in cells. *J. Virol.* **87**, 7155–7159.
- Cimarelli, A., and Luban, J. (1999). Translation elongation factor 1- α interacts specifically with the human immunodeficiency virus type 1 Gag polyprotein. *J. Virol.* **73**, 5388–5401.
- Clavel, F., and Orenstein, J.M. (1990). A mutant of human immunodeficiency virus with reduced RNA packaging and abnormal particle morphology. *J. Virol.* **64**, 5230–5234.
- Clever, J.L., and Parslow, T.G. (1997). Mutant human immunodeficiency virus type 1 genomes with defects in RNA dimerization or encapsidation. *J. Virol.* **71**, 3407–3414.
- Clever, J., Sasseti, C., and Parslow, T.G. (1995). RNA secondary structure and binding sites for gag gene products in the 5' packaging signal of human immunodeficiency virus type 1. *J. Virol.* **69**, 2101–2109.
- Corcoran, D.L., Georgiev, S., Mukherjee, N., Gottwein, E., Skalsky, R.L., Keene, J.D., and Ohler, U. (2011). PARalyzer: definition of RNA binding sites from PAR-CLIP short-read sequence data. *Genome Biol.* **12**, R79.
- Das, A.T., Klaver, B., Klasens, B.I., van Wamel, J.L., and Berkhout, B. (1997). A conserved hairpin motif in the R-U5 region of the human immunodeficiency virus type 1 RNA genome is essential for replication. *J. Virol.* **71**, 2346–2356.
- De Guzman, R.N., Wu, Z.R., Stalling, C.C., Pappalardo, L., Borer, P.N., and Summers, M.F. (1998). Structure of the HIV-1 nucleocapsid protein bound to the SL3 psi-RNA recognition element. *Science* **279**, 384–388.
- Dick, R.A., Kamynina, E., and Vogt, V.M. (2013). Effect of multimerization on membrane association of Rous sarcoma virus and HIV-1 matrix domain proteins. *J. Virol.* **87**, 13598–13608.
- Fisher, R.J., Rein, A., Fivash, M., Urbaneja, M.A., Casas-Finet, J.R., Medaglia, M., and Henderson, L.E. (1998). Sequence-specific binding of human immunodeficiency virus type 1 nucleocapsid protein to short oligonucleotides. *J. Virol.* **72**, 1902–1909.
- Georgiev, S., Boyle, A.P., Jayasurya, K., Ding, X., Mukherjee, S., and Ohler, U. (2010). Evidence-ranked motif identification. *Genome Biol.* **11**, R19.
- Gorelick, R.J., Nigida, S.M., Jr., Bess, J.W., Jr., Arthur, L.O., Henderson, L.E., and Rein, A. (1990). Noninfectious human immunodeficiency virus type 1 mutants deficient in genomic RNA. *J. Virol.* **64**, 3207–3211.
- Grantham, P., and Perrin, P. (1986). AIDS virus and HTLV-I differ in codon choices. *Nature* **319**, 727–728.
- Hafner, M., Landthaler, M., Burger, L., Khorshid, M., Hausser, J., Berninger, P., Rothballer, A., Ascano, M., Jr., Jungkamp, A.C., Munschauer, M., et al. (2010). Transcriptome-wide identification of RNA-binding protein and microRNA target sites by PAR-CLIP. *Cell* **141**, 129–141.
- Harrison, G.P., and Lever, A.M. (1992). The human immunodeficiency virus type 1 packaging signal and major splice donor region have a conserved stable secondary structure. *J. Virol.* **66**, 4144–4153.
- Hill, C.P., Worthylake, D., Bancroft, D.P., Christensen, A.M., and Sundquist, W.I. (1996). Crystal structures of the trimeric human immunodeficiency virus type 1 matrix protein: implications for membrane association and assembly. *Proc. Natl. Acad. Sci. USA* **93**, 3099–3104.
- Houzet, L., Paillart, J.C., Smagulova, F., Maurel, S., Morichaud, Z., Marquet, R., and Mouguel, M. (2007). HIV controls the selective packaging of genomic, spliced viral and cellular RNAs into virions through different mechanisms. *Nucleic Acids Res.* **35**, 2695–2704.
- Jouvenet, N., Simon, S.M., and Bieniasz, P.D. (2009). Imaging the interaction of HIV-1 genomes and Gag during assembly of individual viral particles. *Proc. Natl. Acad. Sci. USA* **106**, 19114–19119.
- Kaye, J.F., Richardson, J.H., and Lever, A.M. (1995). cis-acting sequences involved in human immunodeficiency virus type 1 RNA packaging. *J. Virol.* **69**, 6588–6592.
- Kutluay, S.B., and Bieniasz, P.D. (2010). Analysis of the initiating events in HIV-1 particle assembly and genome packaging. *PLoS Pathog.* **6**, e1001200.
- Kuzembayeva, M., Dilley, K., Sardo, L., and Hu, W.S. (2014). Life of psi: how full-length HIV-1 RNAs become packaged genomes in the viral particles. *Virology* **454–455**, 362–370.
- Kypr, J., and Mrázek, J. (1987). Unusual codon usage of HIV. *Nature* **327**, 20.
- Laham-Karam, N., and Bacharach, E. (2007). Transduction of human immunodeficiency virus type 1 vectors lacking encapsidation and dimerization signals. *J. Virol.* **81**, 10687–10698.
- Lever, A., Gottlinger, H., Haseltine, W., and Sodroski, J. (1989). Identification of a sequence required for efficient packaging of human immunodeficiency virus type 1 RNA into virions. *J. Virol.* **63**, 4085–4087.
- Levin, J.G., Mitra, M., Mascarenhas, A., and Musier-Forsyth, K. (2010). Role of HIV-1 nucleocapsid protein in HIV-1 reverse transcription. *RNA Biol.* **7**, 754–774.
- Licatalosi, D.D., Mele, A., Fak, J.J., Ule, J., Kayikci, M., Chi, S.W., Clark, T.A., Schweitzer, A.C., Blume, J.E., Wang, X., et al. (2008). HITS-CLIP yields genome-wide insights into brain alternative RNA processing. *Nature* **456**, 464–469.
- Lu, K., Heng, X., Garyu, L., Monti, S., Garcia, E.L., Kharytonchyk, S., Dorjsuren, B., Kulandaivel, G., Jones, S., Hiremath, A., et al. (2011a). NMR detection of structures in the HIV-1 5'-leader RNA that regulate genome packaging. *Science* **334**, 242–245.
- Lu, K., Heng, X., and Summers, M.F. (2011b). Structural determinants and mechanism of HIV-1 genome packaging. *J. Mol. Biol.* **410**, 609–633.
- Luban, J., and Goff, S.P. (1994). Mutational analysis of cis-acting packaging signals in human immunodeficiency virus type 1 RNA. *J. Virol.* **68**, 3784–3793.
- Malim, M.H., Tiley, L.S., McCarn, D.F., Rusche, J.R., Hauber, J., and Cullen, B.R. (1990). HIV-1 structural gene expression requires binding of the Rev trans-activator to its RNA target sequence. *Cell* **60**, 675–683.
- McBride, M.S., and Panganiban, A.T. (1996). The human immunodeficiency virus type 1 encapsidation site is a multipartite RNA element composed of functional hairpin structures. *J. Virol.* **70**, 2963–2973.
- McBride, M.S., and Panganiban, A.T. (1997). Position dependence of functional hairpins important for human immunodeficiency virus type 1 RNA encapsidation in vivo. *J. Virol.* **71**, 2050–2058.
- McBride, M.S., Schwartz, M.D., and Panganiban, A.T. (1997). Efficient encapsidation of human immunodeficiency virus type 1 vectors and further characterization of cis elements required for encapsidation. *J. Virol.* **71**, 4544–4554.
- Muriaux, D., Mirro, J., Harvin, D., and Rein, A. (2001). RNA is a structural element in retrovirus particles. *Proc. Natl. Acad. Sci. USA* **98**, 5246–5251.
- Onafuwa-Nuga, A.A., Telesnitsky, A., and King, S.R. (2006). 7SL RNA, but not the 54-kd signal recognition particle protein, is an abundant component of both infectious HIV-1 and minimal virus-like particles. *RNA* **12**, 542–546.
- Ott, D.E., Coren, L.V., and Gagliardi, T.D. (2005). Redundant roles for nucleocapsid and matrix RNA-binding sequences in human immunodeficiency virus type 1 assembly. *J. Virol.* **79**, 13839–13847.
- Ramalingam, D., Duclair, S., Datta, S.A., Ellington, A., Rein, A., and Prasad, V.R. (2011). RNA aptamers directed to human immunodeficiency virus type 1 Gag polyprotein bind to the matrix and nucleocapsid domains and inhibit virus production. *J. Virol.* **85**, 305–314.
- Rein, A., Datta, S.A., Jones, C.P., and Musier-Forsyth, K. (2011). Diverse interactions of retroviral Gag proteins with RNAs. *Trends Biochem. Sci.* **36**, 373–380.
- Richardson, J.H., Child, L.A., and Lever, A.M. (1993). Packaging of human immunodeficiency virus type 1 RNA requires cis-acting sequences outside the 5' leader region. *J. Virol.* **67**, 3997–4005.
- Rulli, S.J., Jr., Hibbert, C.S., Mirro, J., Pederson, T., Biswal, S., and Rein, A. (2007). Selective and nonselective packaging of cellular RNAs in retrovirus particles. *J. Virol.* **81**, 6623–6631.
- Russell, R.S., Hu, J., Bériault, V., Moulard, A.J., Laughrea, M., Kleiman, L., Wainberg, M.A., and Liang, C. (2003). Sequences downstream of the 5' splice

donor site are required for both packaging and dimerization of human immunodeficiency virus type 1 RNA. *J. Virol.* 77, 84–96.

Saad, J.S., Miller, J., Tai, J., Kim, A., Ghanam, R.H., and Summers, M.F. (2006). Structural basis for targeting HIV-1 Gag proteins to the plasma membrane for virus assembly. *Proc. Natl. Acad. Sci. USA* 103, 11364–11369.

Sharp, P.M. (1986). What can AIDS virus codon usage tell us? *Nature* 324, 114.

Shkriabai, N., Datta, S.A., Zhao, Z., Hess, S., Rein, A., and Kvaratskhelia, M. (2006). Interactions of HIV-1 Gag with assembly cofactors. *Biochemistry* 45, 4077–4083.

Sundquist, W.I., and Kräusslich, H.G. (2012). HIV-1 assembly, budding, and maturation. *Cold Spring Harb. Perspect. Med.* 2, a006924.

Wright, E.R., Schooler, J.B., Ding, H.J., Kieffer, C., Fillmore, C., Sundquist, W.I., and Jensen, G.J. (2007). Electron cryotomography of immature HIV-1 virions reveals the structure of the CA and SP1 Gag shells. *EMBO J.* 26, 2218–2226.

Zhou, W., Parent, L.J., Wills, J.W., and Resh, M.D. (1994). Identification of a membrane-binding domain within the amino-terminal region of human immunodeficiency virus type 1 Gag protein which interacts with acidic phospholipids. *J. Virol.* 68, 2556–2569.

lncRNA Directs Cooperative Epigenetic Regulation Downstream of Chemokine Signals

Zhen Xing,^{1,7} Aifu Lin,^{1,7} Chunlai Li,^{1,7} Ke Liang,^{1,7} Shouyu Wang,¹ Yang Liu,^{1,8} Peter K. Park,¹ Li Qin,⁵ Yongkun Wei,¹ David H. Hawke,⁴ Mien-Chie Hung,^{1,2,6} Chunru Lin,^{1,2,*} and Liuqing Yang^{1,2,3,*}

¹Department of Molecular and Cellular Oncology

²Cancer Biology Program, The University of Texas Graduate School of Biomedical Sciences

³Center for RNA Interference and Non-Coding RNAs

⁴Department of Systems Biology

The University of Texas MD Anderson Cancer Center, Houston, TX 77030, USA

⁵Department of Molecular and Cellular Biology, Baylor College of Medicine, Houston, TX 77030, USA

⁶Graduate Institute of Cancer Biology and Center for Molecular Medicine, China Medical University, Taichung 404, Taiwan

⁷Co-first author

⁸Present address: The Institute of Immunology, Zhejiang University School of Medicine, 310058 Hangzhou, China

*Correspondence: clin2@mdanderson.org (C.L.), lyang7@mdanderson.org (L.Y.)

<http://dx.doi.org/10.1016/j.cell.2014.10.013>

SUMMARY

lncRNAs are known to regulate a number of different developmental and tumorigenic processes. Here, we report a role for lncRNA *BCAR4* in breast cancer metastasis that is mediated by chemokine-induced binding of *BCAR4* to two transcription factors with extended regulatory consequences. *BCAR4* binding of SNIP1 and PNUTS in response to CCL21 releases the SNIP1's inhibition of p300-dependent histone acetylation, which in turn enables the *BCAR4*-recruited PNUTS to bind H3K18ac and relieve inhibition of RNA Pol II via activation of the PP1 phosphatase. This mechanism activates a noncanonical Hedgehog/GLI2 transcriptional program that promotes cell migration. *BCAR4* expression correlates with advanced breast cancers, and therapeutic delivery of locked nucleic acids (LNAs) targeting *BCAR4* strongly suppresses breast cancer metastasis in mouse models. The findings reveal a disease-relevant lncRNA mechanism consisting of both direct coordinated protein recruitment and indirect regulation of transcription factors.

INTRODUCTION

Emerging evidence has purported long noncoding RNA (lncRNA) as a new class of players involved in the development and progression of cancer (Fatica and Bozzoni, 2014). However, the regulatory roles played by lncRNAs in breast-cancer-associated aberrant signaling pathways/transcriptional programs are not completely understood. lncRNAs exert their regulatory functions through specific interactions with proteins, including via epigenetic modifiers, transcriptional factors/coactivators, and RNP complexes (Rinn and Chang, 2012). The specific lncRNA-protein interactions could be mediated by canonical RNA-binding domains (RBDs) (Lunde et al., 2007) or by noncanonical

RBDs, including tryptophan-aspartic acid 40 (WD40) domain and RNA-binding domain abundant in Apicomplexans (RAP), demonstrated by recent mRNA interactome capture methodology (Castello et al., 2012). Therefore, it is of great interest to uncover new functions of lncRNAs by dissecting lncRNA-protein interactions mediated by noncanonical RBDs in certain biological processes.

The aberrant activation of the hedgehog signaling pathway in breast cancer has been connected with increased expression of the transcription factor, glioma-associated oncogene homolog 1/2 (GLI1/2) (ten Haaf et al., 2009). GLI1/2-dependent target gene transcription has been shown to be involved in tumor cell growth and metastasis in solid tumors (Rubin and de Sauvage, 2006). However, GLI-target transcription might be activated in the absence of the hedgehog ligand sonic hedgehog (SHH), especially in triple-negative breast cancer (TNBC) (Hui et al., 2013), suggesting that other mechanisms/regulators may regulate the activity of the GLI transcription factor. The direct binding of lncRNAs to transcription factors (Geisler and Collier, 2013) led us to speculate that the association of transcription factor GLI with lncRNAs may function in regulating GLI-dependent transcriptional program essential for breast cancer progression and metastasis.

The lncRNAs implicated in breast cancer represent a promising class of therapeutic targets. Targeting noncoding RNAs by using locked nucleic acid (LNA)-based antisense oligonucleotides strategy has been a longstanding interest (Dias and Stein, 2002), with several successful applications in targeting miRNAs in cancer (Ling et al., 2013). However, therapeutic targeting of lncRNA has not been well documented for breast cancer. Thus, we aimed to determine the therapeutic potential of targeting breast cancer-upregulated lncRNAs by a LNA-based antisense oligonucleotides strategy.

Here, we report the identification of a signaling pathway that is triggered by CCL21 and mediated by citron (rho-interacting, serine/threonine kinase 21) (CIT) kinase to phosphorylate the transcriptional factor GLI2, which regulates target gene expression in breast cancer cells. The lncRNA *BCAR4* is required for phospho-GLI2-dependent gene activation via its direct

interaction with Smad nuclear-interacting protein 1 (SNIP1) and Serine/threonine-protein phosphatase 1 regulatory subunit 10 (PPP1R10, also known as PNUTS). Mechanistically, the *BCAR4*-SNIP1 binding releases the inhibitory role of SNIP1 on p300 histone acetyltransferase (HAT) activity, leading to the acetylation of histones, including a novel mark, H3K18ac, on the promoters of *GLI2* target transcription units. The acetylated H3K18 can be further recognized by PNUTS, which is recruited to the promoters of *GLI2* target genes by *BCAR4*, to attenuate the protein's inhibitory effect on the enzymatic activity of PP1, leading to hypophosphorylation of RNA polymerase II at Ser5. Elevated *BCAR4* expression is correlated with higher metastatic potential and shorter survival time of breast cancer patients, whereas its therapeutic inhibition by LNA displays *in vivo* efficacy against metastasis. Our findings have provided supporting evidence for the regulatory roles played by lncRNAs in the progression of aggressive breast cancers. Broadly, our results of the therapeutic effectiveness of *BCAR4* LNA against breast cancer metastasis document an example showing the pharmacologic value of lncRNA in human cancer and other diseases.

RESULTS

BCAR4 Correlates with Advanced Breast Cancer and Regulates GLI-Mediated Transcription

To identify breast cancer-relevant lncRNAs, we profiled the expression of lncRNAs in two stage III breast cancer tissues and their paired adjacent noncancerous tissues (Figure S1A available online) by lncRNA Array (v. 3.0) (ArrayStar). An average of 1,381 upregulated lncRNAs (range from 1,034 to 1,729) and 1,458 downregulated lncRNAs (range 1,408–1,508) with significantly differential expression (≥ 3.0 -fold) were identified (Figure 1A; Table S1). We further compared the lncRNA expression levels between breast cancer tissues and their paired adjacent normal tissues based on the NCBI RefSeq database (which contains 3,991 human lncRNAs with annotated NR accession number), identifying 65 and 116 upregulated lncRNAs in two patient cases, respectively (≥ 4.0 -fold) (Figure 1B). Among these lncRNAs, 21 were consistently upregulated in both patient samples, of which *BCAR4*, initially identified through genetic screening as a novel gene involved in tamoxifen resistance in breast cancers (Meijer et al., 2006), showed the most upregulation (logFC: 15.9 and 16.1, respectively) (Figures S1B and S1C).

We first performed RNA *in situ* hybridization on breast cancer tissue microarrays (clinicopathological features listed in Table S2) using RNAScope 2.0 HD technology to examine the potential correlation of *BCAR4* with breast cancer. In a training set of breast cancer tissue microarrays containing 232 cases, *BCAR4* exhibited positive staining only in 10% of the normal breast tissues, whereas 54.10% of breast cancer tissues showed positive *BCAR4* expression ($p = 0.0057$) (Figure 1C). In a validation set containing 170 cases, none of ten normal adjacent breast tissues showed detectable *BCAR4* expression, but 61.88% of breast cancer tissues exhibited positive *BCAR4* staining ($p = 0.0011$) (Figure 1C). Furthermore, breast cancer at advanced lymph node metastasis stage (TnN > 0M \geq 0) showed increased *BCAR4* expression compared to those at early-stage tumor with no lymph node metastasis (TnN0M0) ($p = 0.0001$,

training set; $p = 0.0035$, validation set) (Figure 1D). Elevated *BCAR4* expression also significantly correlates with shorter survival time of breast cancer patients ($n = 160$; $p = 0.0145$) (Figure 1E). We further analyzed breast cancer database in Oncomine, finding that *BCAR4* expression not only correlates with breast cancer but also with triple negativity lymph node metastasis and 5 years recurrence (Figure S1D). Oncomine database also showed significant correlation of *BCAR4* expression with metastatic prostate cancer, lung cancer, colorectal cancer, and rectal cancer (Figure S1D). To confirm this, we employed the RNAScope Assay to analyze *BCAR4* expression in normal and cancer tissues from multiple organ, observing increased *BCAR4* expression in many types of human cancer tissues, including colorectal, melanoma, and lung cancer, compared to normal tissues (Figure 1F; Table S3). Taken together, these results demonstrated the strong correlation of *BCAR4* expression with breast cancer progression and the relevance of elevated *BCAR4* expression to human cancer development and progression.

We then examined the expression of *BCAR4* in a panel of breast cancer cell lines, finding higher expression of *BCAR4* in mesenchymal-like cell lines with metastasis potential compared to epithelial-like cell lines, which are considered as nonmetastatic (Figure 1G). Next, we examined the subcellular localization of *BCAR4* by RNA fluorescence *in situ* hybridization (FISH) and qRT-PCR analyses on fractionated RNA, finding that the *BCAR4* transcript is predominately localized in the nucleus (Figures 1H and S1E). *BCAR4* has two major splice variants, full-length transcript (~ 1.3 kb) and an isoform lacking two alternate exons (~ 680 bp), and our northern blot analysis revealed that the full-length isoform was predominately expressed in MDA-MB-231 cells, but truncated isoform barely expressed (Figure S1F). Because the previous report suggested that *BCAR4* may encode a small peptide in bovine oocytes (Th  lie et al., 2007), we generated an antibody using the predicted translated peptide sequence. However, neither immunoblotting of MDA-MB-231 lysate nor *in vitro* translation assays showed protein coding potential of *BCAR4* (Figure S1G; data not shown).

We next analyzed the effect of *BCAR4* knockdown on activation of key signaling pathways in breast cancer cells using the Signal Finder 45-Pathway Reporter Array, finding that either small interfering RNA (siRNA) or LNA efficiently depleted *BCAR4* expression (Figures S1H and S1I) and knockdown of *BCAR4* dramatically inhibited GLI reporter luciferase activity but no other transcription factor reporters (Figure 1I). qRT-PCR analysis confirmed decreased expression of endogenous GLI target genes with *BCAR4* knockdown (Figure 1J). These data suggest the potential role of *BCAR4* in mediating the GLI-dependent hedgehog signaling pathway in breast cancer cells.

Identification and Biochemical Characterization of *BCAR4*-Associated Proteins

Through RNA pull-down, followed by mass spectrometry (MS) analysis, we identified that *in vitro*-transcribed biotinylated *BCAR4* sense transcript associated specifically with CIT kinase, *GLI2*, SNIP1, and PNUTS, even under high stringency wash conditions. However, the antisense transcript of *BCAR4* associated with some general RNA-binding proteins that were also bound

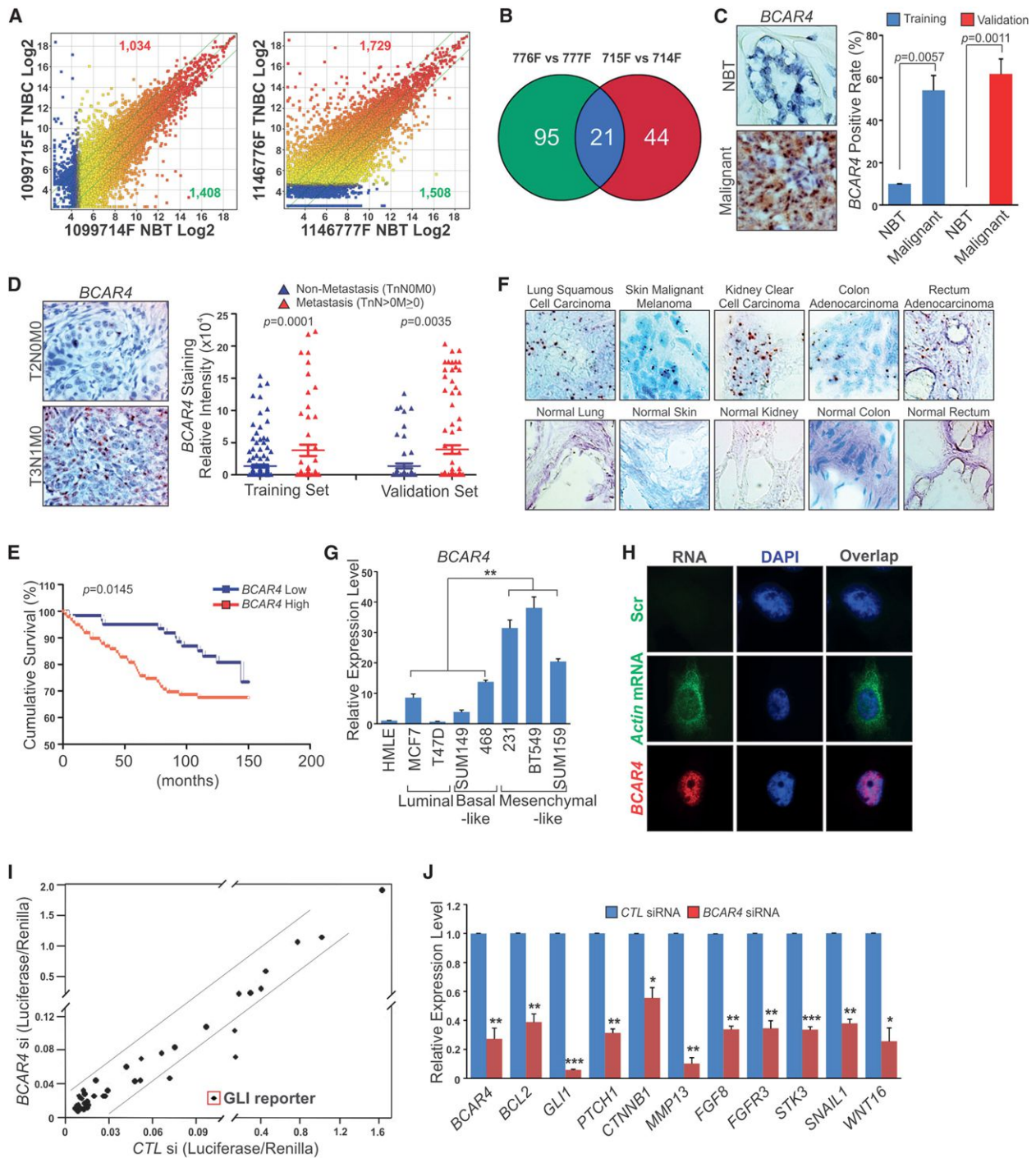


Figure 1. *BCAR4* Correlates with Breast Cancer Metastasis

(A) Scatter plots of lncRNAs significantly upregulated (red) or downregulated (green) in two pairs of TNBC tissues compared to the matched adjacent normal tissues (NBT). x and y axes, averaged normalized signal values (log₂ scaled); green lines, fold changes = 4.

(B) Commonly upregulated lncRNAs in two pairs of TNBC compared to NBT.

(C) RNAScope detection of *BCAR4* expression in human breast cancer and adjacent normal tissues. Left panel: representative images; right panel: statistical analysis of training set (10 normal tissues versus 222 cancer tissues) and validation set (10 normal tissues versus 160 cancer tissues).

(D) RNAScope detection of *BCAR4* expression in nonmetastasis (TnN0M0) versus metastasis (TnN > 0M ≥ 0) breast cancer tissue. Left panel: representative images; right panel: statistical analysis of training set (167 nonmetastasis versus 55 Metastasis) and validation set (66 nonmetastasis versus 94 metastasis).

(E) Kaplan-Meier survival analysis of *BCAR4* expression in breast cancer patients (n = 160).

(F) RNAScope detection of *BCAR4* expression in multiple human tissues.

(legend continued on next page)

by the beads (Figures 2A and S2A; Table S4). Of note, in one of two biological repeats of RNA pull-down experiment, we observed the relative abundant association of *BCAR4* with heterogeneous nuclear ribonucleoprotein, which have been reported to bind other lncRNAs (Carpenter et al., 2013; Huarte et al., 2010). Furthermore, the MS data indicated the potential phosphorylation of GLI2 at Serine149 (Figure S2B).

The RNA pull-down assays with cell lysate further confirmed the specific association of *BCAR4* with the proteins identified by MS analysis (Figure 2B). In vitro RNA-protein binding assay revealed that only PNUTS and SNIP1 directly interact with *BCAR4* (Figures 2C and S2C). Protein domain mapping studies demonstrated that *BCAR4* binds the 97–274 amino acid (aa) region of SNIP1 and 674–750 aa region of PNUTS, respectively (Figures 2D and 2E). The 97–274 aa region of SNIP1 encodes a domain known as the domain of unknown function (DUF) and has been suggested to bind miRNA (Yu et al., 2008), an observation that is consistent with our observation that the DUF of SNIP1 serves as the RNA-binding domain for *BCAR4*. PNUTS also has an RNA-binding motif, the 674–750 aa region known as RGG-box (Kim et al., 2003). To further understand the *BCAR4*-protein interactions in vivo, we performed immunoprecipitation using antibodies against CIT, GLI2, SNIP1, and PNUTS, respectively, under the condition of *BCAR4* knockdown (Figures S2D and S2E), finding that knockdown of *BCAR4* impaired the interaction of PNUTS with proteins CIT, GLI2, and SNIP1, but had minimal effect on the association of CIT, GLI2, and SNIP1 with each other (Figure S2E). Given the observation that only SNIP1 and PNUTS directly bound to *BCAR4* (see Figure 2C), our data suggest that SNIP1 mediates the association of CIT and GLI2 with *BCAR4* and that SNIP1 and PNUTS bind distinct regions of *BCAR4*.

To map the *BCAR4* sequence motifs responsible for SNIP1 and PNUTS binding, we performed an in vitro RNA pull-down followed by dot blot assay (Yang et al., 2013). The motif sequence of *BCAR4* bound/protected by SNIP1 and PNUTS was identified to encompass ²³⁵TGT...GGA²⁸⁸ and ⁹⁹¹GTT...ATA¹⁰⁴⁴, respectively (Figure 2F). However, the glutathione S-transferase (GST) protein showed no specific binding to any region of *BCAR4* (Figure 2F). Deletion of the corresponding sequence of *BCAR4* (Δ 212– Δ 311) abolished its interaction between SNIP1 with no effect on PNUTS binding (Figure 2G). Deletion of the motif sequence Δ 968– Δ 1087 of *BCAR4* abolished its interaction with PNUTS, but not with SNIP1 (Figure 2G). Electrophoretic mobility shift assays (EMSAs) were further used to confirm the direct binding of *BCAR4* with SNIP1 and PNUTS. Incubation of the *BCAR4* RNA probe (nucleotide [nt] 235–288) and (nt 991–1044) with recombinant SNIP1 and PNUTS, respectively, resulted in specific gel retardation (Figure 2H). Under these conditions, no shift was observed when the corresponding cold probes were used (Figure 2H). We, therefore, conclude that *BCAR4* directly bind to SNIP1 and PNUTS via two distinct regions.

Given MS data showing that GLI2 is phosphorylated at Ser149 and associates with CIT kinase (see Figures 2A and S2B), we reasoned that CIT may serve as a kinase to phosphorylate GLI2. An in vitro kinase assay indicated that bacterially expressed wild-type (WT) GLI2 was phosphorylated by CIT, but not S149A mutant (Figure S2F). ULK3 served as the positive control due to its reported ability to phosphorylate GLI (Maloverjan et al., 2010). An in vitro RNA-protein binding assay using biotinylated *BCAR4* and GLI2 proteins phosphorylated by CIT in vitro showed no interaction (Figure S2G).

To investigate the role of GLI2 Ser149 phosphorylation in vivo, we generated rabbit polyclonal antibodies that specifically recognized Ser149-phosphorylated GLI2 referred to as p-GLI2 (Ser149) antibody, which specifically detected bacterially purified GLI2 protein that phosphorylated by CIT in vitro, with minimal reactivity toward GLI2 phosphorylated by ULK3 (Figure 2I). We conclude that p-GLI2 (Ser149) antibody specifically recognizes CIT-mediated Ser149 phosphorylation of GLI2. Next, we evaluate the level of phospho-GLI2 in breast cancer by immunohistochemistry (IHC) analysis of clinical tumor specimens, finding higher p-GLI2 (Ser149) levels in invasive breast cancer tissues compared with adjacent normal tissues ($p = 0.0087$) (Figure 2J). Our IHC staining further revealed increased p-GLI2 (Ser149) level in multiple cancer types compared to their corresponding normal tissues (Figure S2H; Table S5). IHC analysis also revealed higher CIT expression in invasive breast cancer compared with adjacent normal breast tissues ($p = 0.0055$) (Figure S2I) and the staining of phosphorylated GLI2 strongly correlated with that of *BCAR4* and CIT staining (data not shown). Taken together, we identified and characterized that *BCAR4* binds a protein complex containing SNIP1, PNUTS, phosphorylated GLI2, and CIT via its direct interaction with SNIP1 and PNUTS.

CCL21 Induces GLI2 Ser149 Phosphorylation and Nuclear Translocation of Phosphorylated GLI2

The CIT kinase-mediated GLI2 phosphorylation prompted us to investigate whether this phosphorylation could be triggered in MDA-MB-231 cells by hedgehog signaling. Surprisingly, although the ligand SHH activated hedgehog signaling in Daoy cells evidenced by stimulated *SHH* gene induction as previously reported (Wang et al., 2012), minimal effect was observed in MDA-MB-231 cells (Figure S3A), and no phosphorylated GLI2 was detected (data not shown), suggesting that a noncanonical hedgehog signaling pathway, involving Ser149-phosphorylated GLI2, may exist in breast cancer.

We then explored whether extracellular signals that activate CIT kinase could also trigger GLI2 phosphorylation in breast cancer cells. Given that CIT kinase can be activated by GTPase Rho proteins (Madaule et al., 1998), we first screened the CIT-Rho interaction in breast cancer cells. Although CIT kinase is constitutively associated with RhoA as previously reported (Gai et al., 2011), the presence of Rho activator specifically triggered the

(G) qRT-PCR detection of *BCAR4* expression in a panel of cell lines.

(H) Nuclear localization of *BCAR4* detected by RNA FISH in MDA-MB-231 cells.

(I) Identification of signal pathways affected by *BCAR4* knockdown in MDA-MB-231 cells. x and y axes, normalized ratio of firefly/Renilla luciferase activities.

(J) qRT-PCR detection of GLI2-target genes expression. Error bars, SEM of three independent experiments (* $p < 0.05$, ** $p < 0.01$, and *** $p < 0.001$).

See also Figure S1 and Tables S1, S2, and S3.

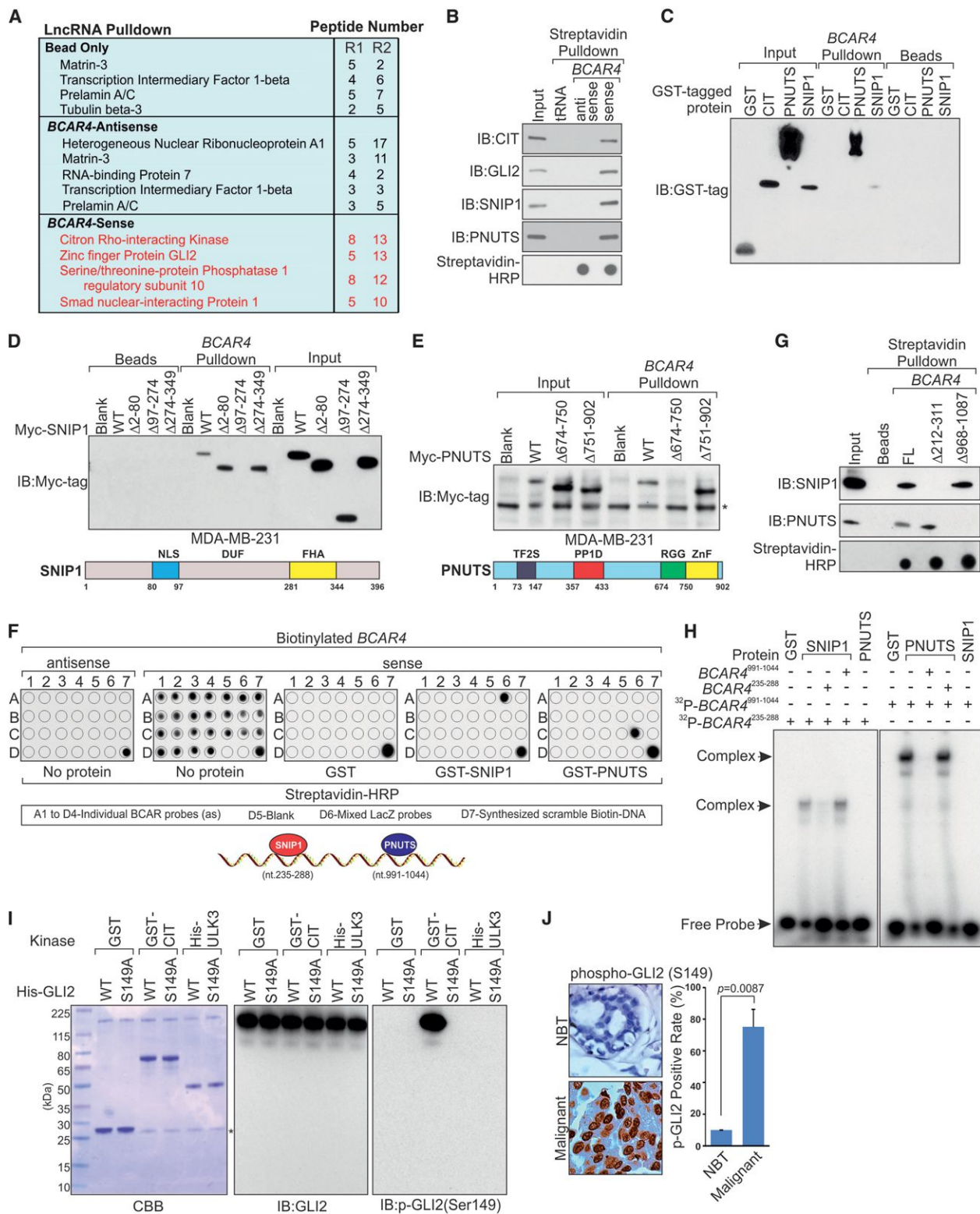


Figure 2. Identification and Biochemical Characterization of *BCAR4*-Associated Proteins

(A) A list of top *BCAR4*-associated proteins identified by RNA pull-down and MS analysis in MDA-MB-231 cells: R1 and R2 (biological repeat 1 and 2).

(B and C) Immunoblot (IB) detection of proteins retrieved by in vitro-transcribed biotinylated *BCAR4* from MDA-MB-231 cell lysates (B) and indicated recombinant proteins (C).

(legend continued on next page)

interaction between RhoC and CIT kinase (Figure S3B). Then we hypothesized that the RhoC-activating stimulus may activate CIT kinase. Indeed, we screened 13 known growth factors/cytokines/chemokine involved in RhoC activation and breast cancer metastasis (Favoni and de Cupis, 2000; Kakinuma and Hwang, 2006), finding that CXCL12, CCL21, IGF-I, PDGF-BB, and TGF- β 1 enhanced the interaction between RhoC and CIT (Figure 3A). The same stimuli induced activation of CIT kinase indicated by phosphorylation of MLC, a classic CIT kinase substrate (Yamashiro et al., 2003), with CCL21 exhibiting the highest induction (Figure S3C). We then tested the phosphorylation of GLI2 in MDA-MB-231 cells treated with CXCL12, CCL21, IGF-1, PDGF-BB, and TGF- β 1, finding that CCL21 dramatically induced Ser149 phosphorylation of GLI2 (Figure 3B), which was significantly reduced by CIT knockdown (Figure 3C). Consistently with previous finding that CCL21-CCR7 autocrine signaling is critical for breast cancer metastasis (Müller et al., 2001), treatment of MDA-MB-231 cells with either neutralizing anti-CCL21 or anti-CCR7 antibodies inhibited basal or CCL21-induced GLI2 phosphorylation (Figures S3D and S3E). CCL21 treatment also dramatically induced GLI2 Ser149 phosphorylation in a panel of additional cancer cell lines, ruling out the possibility of cell line-specific effect (Figure S3F).

Next, we investigated the functional consequence of Ser149 phosphorylation on GLI2. In the cytoplasm, GLI is associated with the suppressor of fused homolog (SUFU), which regulates the cellular localization of GLI (Dunaeva et al., 2003). We performed coimmunoprecipitation experiments and observed that CCL21 treatment induced dissociation between GLI2 and SUFU (Figure S3G), whereas the exogenously expressed GLI2 S149A mutant failed to release from SUFU in response to CCL21 (Figure 3D). Given that SNIP1, which is in the same complex with GLI2 (see Figure 2A), harbors an FHA domain that recognizes phosphoserine/threonine, we hypothesized that Ser149 phosphorylation of GLI2 is required for its interaction with SNIP1 via the FHA domain. Indeed, either knockdown of CIT or introduction of S149A mutant reduced CCL21-induced interaction between GLI2 and SNIP1 (Figures 3C and 3E). Consistently, deletion or point mutation of amino acids that are critical for FHA domain function (Durocher et al., 2000) also abolished SNIP1's interaction with phosphorylated GLI2 (Figures 3F and 3G). We then performed nuclear fractionation experiments, finding that phosphorylated GLI2 translocated to the nucleus upon CCL21 treatment; whereas CIT, SNIP1, and PNUTS did not exhibit relocation (Figure 3H). The phospho-GLI2-specific antibody also exhibited nuclear staining patterns in breast cancer tissue samples (see Figure 2J). Knockdown of CIT or

SNIP1 abolished CCL21-induced nuclear translocation of GLI2 (Figure 3I). In accordance with this, GLI2 S149A mutant failed to translocate into the nucleus upon CCL21 treatment (Figure S3H). Our findings reveal a CCL21/CIT kinase/phospho-GLI2/SNIP1 signaling cascade in breast cancer cells, which may represent a noncanonical hedgehog pathway.

BCAR4 Is Required for Transcription Activation of Phospho-GLI2-Dependent Target Genes in Breast Cancer Cells

To test if CCL21/CIT/SNIP1 signaling axis-mediated phospho-GLI2 nuclear translocation leads to the activation of GLI target genes, we performed a chromatin immunoprecipitation (ChIP) assay using antibodies against GLI2 or phospho-GLI2, finding that Ser149 phosphorylated GLI2 was present on the promoters of several well-established GLI target genes *PTCH1*, *IL-6*, *MUC5AC*, and *TGF- β 1*, but not on the promoter of a non-GLI target gene, *RPLP0* (Figures 4A and 4B). We then performed a chromatin isolation by RNA purification (ChIRP) assay to examine the genomic occupancy of *BCAR4*, finding that in response to CCL21 treatment, *BCAR4* was recruited to the promoters of *PTCH1*, *IL-6*, *MUC5AC*, and *TGF- β 1* (Figures 4C, S3I, and S3J). Consistently, either knockdown of *BCAR4* or overexpression of GLI2 S149A mutant dramatically impaired CCL21-induced expression of *PTCH1*, *IL-6*, *MUC5AC*, and *TGF- β 1* genes (Figure 4D; data not shown).

One of the major biological roles of GLI is to modulate the gene expression related to cell migration and invasion (Feldmann et al., 2007). Thus, we examined the effect of GLI2, *BCAR4*, and other *BCAR4* bound proteins on breast cancer cell invasion and migration. The treatment of MDA-MB-231 cells with validated siRNAs against *BCAR4*, *CIT*, *SNIP1*, or *PNUTS* or neutralizing antibody against CCL21 all dramatically inhibited cell migration (Figures 4E–4G) and invasion (Figure 4H; data not shown) but did not affect cell proliferation (Figure S4A). Consistently, stable knockdown of *BCAR4* by small hairpin RNAs (shRNAs) in MDA-MB-231 LM2 cells reduced migration and invasion properties of these cells (Figures S4B–S4D). We also tested if *BCAR4* is critical for migration and invasion of those metastatic cancer cell lines that respond to CCL21 treatment (see Figure S3F). Our data showed that while knockdown of *BCAR4* had no effect on proliferation of HCT116, H1299, HepG2, and Hey8 cells (Figures S4E and S4F), the migration and invasion of these cells were significantly reduced (Figures S4G and S4H; data not shown). In addition, CCL21-induced GLI2 target genes expression in these cell lines was inhibited by *BCAR4* knockdown (Figures S4I and S4J; data not shown).

(D and E) IB detection of Myc-tagged SNIP1 (D) and PNUTS (E) (WT versus domain truncation mutants) retrieved by in vitro-transcribed biotinylated *BCAR4*. Lower panels: graphic illustration of the domain structure of SNIP1 (D) or PNUTS (E).

(F) In vitro RNA-protein binding, followed by dot blot assays. Bottom panel: schematic illustration of the *BCAR4* sequence motifs that is recognized by SNIP1 and PNUTS, respectively.

(G) IB detection of proteins retrieved by in vitro-transcribed biotinylated *BCAR4* (WT versus Δ 212– Δ 311 and Δ 968– Δ 1087) from MDA-MB-231 cell lysates.

(H) EMSA of recombinant SNIP1 and PNUTS binding to *BCAR4* nt 235–288 and nt 991–1044, respectively.

(I) In vitro kinase assay showing CIT-mediated phosphorylation of GLI2 (WT versus S149A). *, unspecific band.

(J) IHC staining of phospho-GLI2 (S149) in human breast cancer and adjacent normal tissues. Left: representative image. Right: statistics analysis based on 10 normal tissues versus 222 cancer tissues is shown.

See also Figure S2 and Tables S2, S4, and S5.

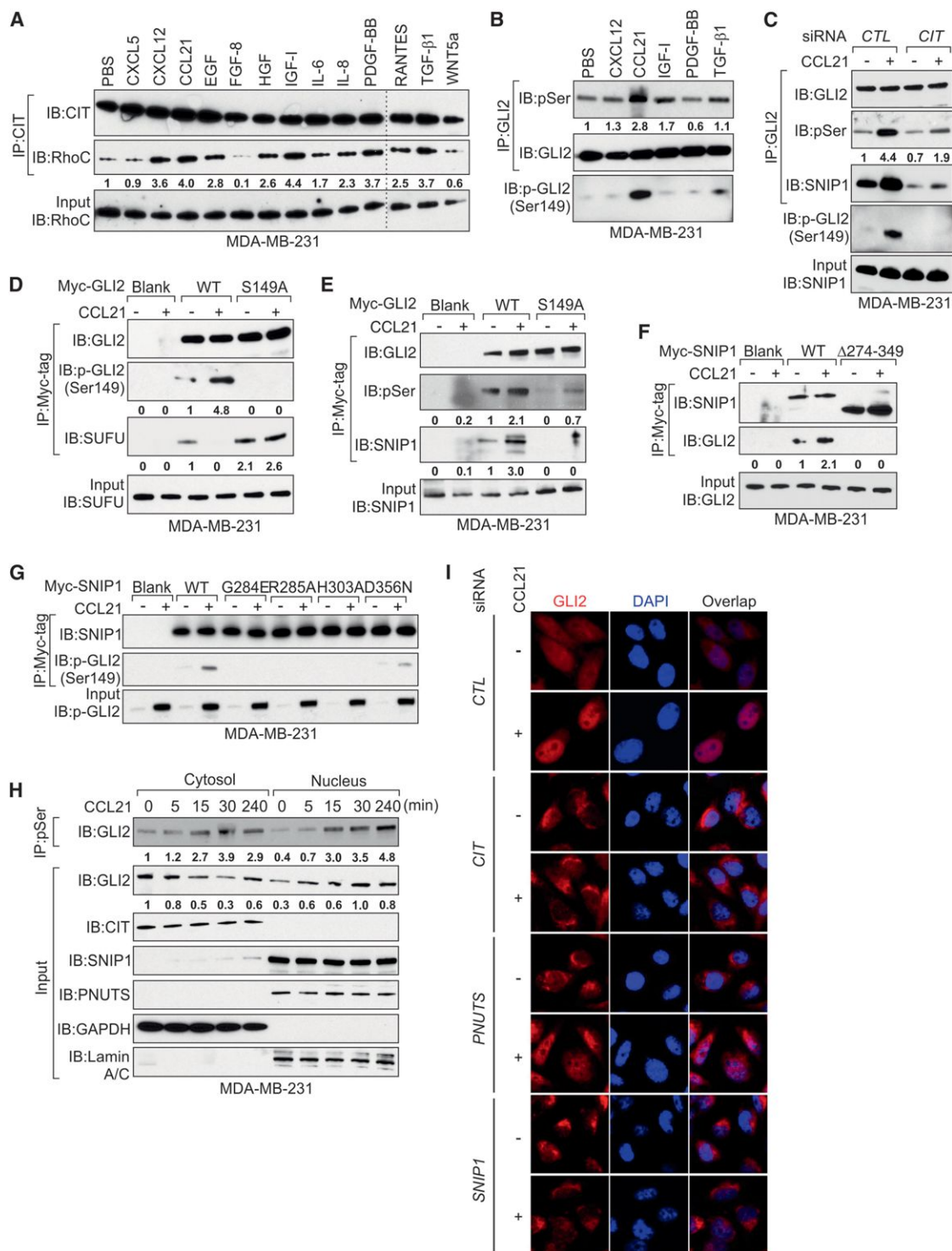


Figure 3. Identification of a Noncanonical Hedgehog Signaling Pathway Mediated by CCL21/CIT/Phospho-GLI2 Signaling Axis

(A and B) Immunoprecipitation (IP) and IB detection of CIT-RhoC interactions (A) and GLI2 phosphorylations (B) in cells treated with indicated growth factors, cytokines, or chemokines.

(C) IP and IB of GLI2 phosphorylations in cells transfected with indicated siRNAs followed by CCL21 treatment.

(D and E) IP and IB detection of GLI2-SUFU (D) or GLI2-SNIP1 (E) interactions in MDA-MB-231 cells transfected with GLI2 (WT versus S149A) followed by CCL21 treatment.

(legend continued on next page)

Given that *BCAR4* is critical for metastasis potential of cancer cells and our observation of lower *BCAR4* expression level in nonmetastatic breast cancer cell lines compared to metastatic breast cell lines (see Figure 1G), we reasoned that overexpression of *BCAR4* in a nonmetastatic cell line may increase its metastasis potential. MCF-7 is a nonmetastatic breast cancer cell line but expresses the CCR7, the receptor for CCL21 (Müller et al., 2001). Indeed, stimulation of MCF-7 cells with CCL21 modestly enhanced their invasion (Figure 4I). However, overexpression of full-length *BCAR4*, but not the deletion mutants abolishing SNIP1 or PNUITS binding in MCF-7 cells (Figure S4K), increased the invasion and GLI2 target genes expression even under the basal condition (Figures 4I, 4J, and S4L), which was not due to cell proliferation effect (Figure S4M). These data strongly argue the important role of *BCAR4* in the phospho-GLI2-mediated transcription activation of a subset of genes, which may contribute to breast cancer cell migration and invasion.

***BCAR4* Binds SNIP1 and Releases the Inhibitory Effect of SNIP1 on p300 HAT Activity**

Next, we investigated the molecular mechanism by which *BCAR4* regulates GLI2 target genes expression. Considering that *BCAR4* directly interacts with SNIP1 *in vitro*, we explored whether this interaction is functionally important *in vivo* by examining the SNIP1-*BCAR4* interaction by an RNA immunoprecipitation (RIP) assay, finding that in response to CCL21 treatment, SNIP1 bound to *BCAR4* in multiple cancer cell lines (Figures S5A–S5C). As a control, no interaction between SNIP1 and *NEAT2*, an abundant nuclear lncRNA, was observed (Figures S5A–S5C). As expected, deletion of the 97–274 aa region abolished SNIP1-*BCAR4* interaction (Figure 5A), which is consistent with our previous observation that the DUF domain of SNIP1 is required for SNIP1-*BCAR4* interaction (see Figure 2D). Surprisingly, deletion of the FHA domain (region 274–349 aa) of SNIP1 led to constitutive SNIP1-*BCAR4* interaction (Figures 5A and S5D), suggesting that binding to phosphoserine/threonine via its FHA domain, is required for SNIP1's subsequent interaction with *BCAR4*, possibly through a mechanism involving the conformational change of SNIP1 upon phospho-GLI2 binding. Indeed, FHA domain mutants of SNIP1 all failed to interact with *BCAR4*, whereas wild-type SNIP1 along with the D356N mutant, which exhibits no effect on phospho-GLI2 binding, was able to bind *BCAR4* (Figure 5B). These data suggest that SNIP1's FHA domain may block the DUF domain, preventing SNIP1-*BCAR4* interaction. Upon stimulation, the FHA domain recognizes phospho-Ser149 of GLI2, which causes conformational changes that may expose the DUF domain for *BCAR4* binding.

SNIP1 has been reported to interact with p300 and potentially regulates p300-dependent gene transcription (Kim et al., 2000). Although immunoprecipitation of SNIP1 confirmed its interaction with p300, the interaction was not affected by deprivation of

BCAR4 (Figure S5E). Deletion of either DUF domain of SNIP1 (region 97–274 aa) or the *BCAR4* SNIP1 binding motif (nt 212–311) exhibited minimal effect on SNIP1-p300 interaction (Figures S5F and S5G). We then examined the HAT activity of p300 in the presence of SNIP1 and/or *BCAR4*. Surprisingly, the HAT activity of p300, was strongly inhibited by recombinant SNIP1 but could be rescued by *in vitro*-transcribed *BCAR4* RNA (Figure 5C). This rescue was dependent on the interaction between *BCAR4* and SNIP1's DUF domain because the presence of *BCAR4* alone had no effect on the HAT activity of p300. Moreover, deletion of *BCAR4*'s SNIP1 binding motif (nt 212–311) abolished the rescue of p300's HAT activity (Figure 5C). Therefore, our data indicated that the interaction between SNIP1 and *BCAR4* released the inhibitory role of SNIP1 on the HAT activity of p300.

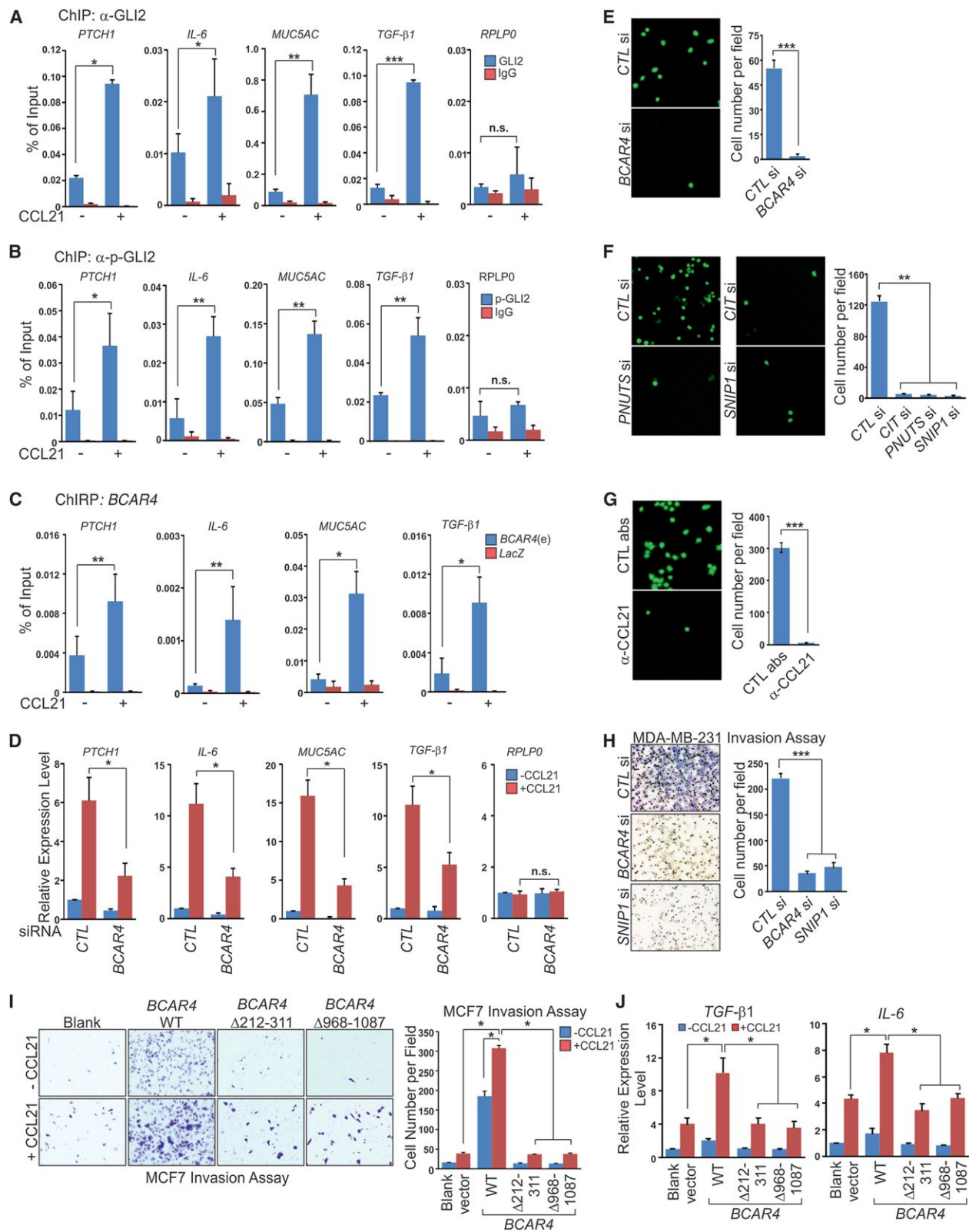
Although it has been suggested that SNIP1 regulates the p300-dependent transcription of multiple signaling pathways (Fujii et al., 2006; Kim et al., 2000, 2001), the mechanism is not clear. We mapped the domains of SNIP1 that may interact with p300 and found that while both the N-terminal (2–80 aa) and DUF domain (97–274 aa) of SNIP1 were required for p300 binding (Figure S5H), the DUF domain of SNIP1 is the minimum region required to inhibit the enzymatic activity of p300 (Figure 5D). By incubating SNIP1 with p300 catalytic unit (aa 1198–1806) and derivative truncation mutants, we found that the DUF domain of SNIP1 interact with PHD (aa 1198–1278) and CH3 domains (aa 1664–1806) of p300 catalytic unit, which may interfere with p300's HAT activity (Figure 5E). According to our *in vitro* observations, the DUF domain also binds *BCAR4*, raising a possible role of *BCAR4* in regulating p300's HAT activity. Indeed, in the presence of BSA and tRNA, p300 exhibited dose-dependent HAT activity, which was abolished in the presence of SNIP1 DUF domain alone (Figure 5F). In contrast, in the presence of sense but not anti-sense *BCAR4*, p300 HAT activity was largely rescued (Figure 5F). These data suggest that the DUF domain of SNIP1 binds PHD and CH3 domains of p300 to inhibit the HAT activity, whereas signal-induced binding of *BCAR4* to SNIP1 DUF domain releases its interaction with the catalytic domain of p300, leading to the activation of p300.

p300-mediated histone acetylation is critical for transcription activation (Wang et al., 2008). We then screened histone acetylation on GLI2 target gene promoters, finding that H3K18ac, H3K27ac, H3K56ac, H4K8ac, H4K12ac, and H4K16ac were induced by CCL21 treatment in breast cancer cells, with H3K18ac showing the highest level (Figure 5G). Knockdown of *BCAR4* abolished CCL21-induced H3K18 acetylation on GLI2 target gene promoters; however, this was not due to reduced recruitment of phosphorylated-GLI2 or p300 to GLI2 (Figure 5H). These findings suggest that *BCAR4* activates p300 by binding SNIP1's DUF domain to release the inhibitory effect of SNIP1 on p300, which results in the acetylation of histone marks required for gene activation.

(F and G) IP and IB detection of GLI2-SNIP1 interactions in cells transfected with SNIP1 (WT versus $\Delta 274$ – $\Delta 349$) (F) or (WT versus FHA domain point mutants) (G) followed by CCL21 treatment.

(H and I) IP and IB (H) and Immunofluorescence (I) detection of phospho-GLI2 nuclear translocation in cells treated with CCL21 treatment at different time points (H) or transfected with indicated siRNAs followed by CCL21 treatment (I).

See also Figure S3.



(legend on next page)

Recognition of *BCAR4*-Dependent Histone Acetylation by PNUTS Attenuates Its Inhibitory Effect on PP1 Activity

Based on our data that the 3' of *BCAR4* interacts with PNUTS in vitro, we next examined this interaction in vivo by RIP experiments. We found that PNUTS constitutively interacts with *BCAR4* via its RGG domain (Figures 6A, S5A–S5C, and S6A), an observation that is consistent with our in vitro data (see Figure 2E). PNUTS functions as a regulatory subunit for PP1, inhibiting the phosphatase activity of PP1 (Kim et al., 2003). As such, we wondered whether *BCAR4* could regulate PP1's phosphatase activity via binding PNUTS. The immunoprecipitation assay indicated that knockdown of *BCAR4* has minimal effect on PNUTS-PP1A interaction (Figures S1I and S6B). As previously reported (Kim et al., 2003), the phosphatase activity of PP1 was inhibited by PNUTS (Figure S6C). However, neither sense nor antisense *BCAR4* could rescue PP1's activity (Figure S6D), leading us to explore whether any histone modifications could rescue PP1 activity given that recruitment of the PNUTS/PP1 complex by *BCAR4* could possibly activate the transcription of GLI2 target genes.

Surprisingly, the inhibition of PP1's phosphatase activity by PNUTS was largely rescued by purified nucleosome from HeLa cells but not by recombinant nucleosome, whereas neither nucleosome alone affected PP1 activity (Figure 6B), suggesting that modified histones binding is critical to release PNUTS's inhibitory effect on PP1 activity. We then utilized a modified histone peptide array to test this possibility, finding that PNUTS, but not SNIP1, directly recognized acetylated histones, including H4K20ac, H3K18ac, H3K9ac, H3K27ac, and H4K16ac (Figure 6C), which was confirmed by histone peptide pull-down experiments (Figure 6D). A previous study indicated that a minimum region from 445–450 aa of PNUTS is required to inhibit the phosphatase activity of PP1 (Kim et al., 2003). We then examined if acetylated histone could also recognize this region, finding that deletion of aa 443–455 of PNUTS abolished its interaction with acetylated histone H3 (Figure 6E), suggesting that the inhibitory role of PNUTS, mediated by motif aa 443–455, is attenuated in the presence of acetylated histone, leading to activation of PP1 enzymatic activity. Consistently, acetylated, but not methylated, histone peptides specifically rescued PP1 activity from PNUTS inhibition (Figure 6F).

PP1 has been reported to dephosphorylate the carboxyl terminal domain (CTD domain) of RNA polymerase II at Ser5, which is accumulated at promoter regions of target genes (Komarnitsky et al., 2000; Washington et al., 2002). A recent study showed that depletion of PNUTS in *Drosophila* results in global hyperphosphorylation of RNA Pol II Ser5, leading to global tran-

scription pause and development defect (Ciurciu et al., 2013). Therefore, we tested if PNUTS/PP1 regulates phosphorylation of RNA Pol II Ser5, finding that knockdown of PNUTS led to the hyperphosphorylation of RNA Pol II Ser5 (Figures S6E and S6F). We then investigated the functional roles of PNUTS-acetylated histone interaction in regulating the status of RNA Pol II Ser5 phosphorylation in the presence of a p300 inhibitor, C646, which eliminated the histone acetylation as represented by H3K18ac (Figures 6G, S6G, and S6H). Our data indicate that CCL21-triggered recruitment of PNUTS and PP1 to the promoters of GLI2 target genes was not affected by p300 inhibitor (Figures 6G, S6G, and S6H) and that the levels of Pol II Ser5 phosphorylation on these promoters were decreased upon CCL21 treatment (Figures 6G, S6G, and S6H). However, the CCL21-induced hypophosphorylation of RNA Pol II Ser5 was abolished in the presence of the p300 inhibitor (Figures 6G, S6G, and S6H), suggesting that histone acetylation-dependent PP1 activity modulates RNA Pol II Ser5 phosphorylation levels at gene promoter regions. Taken together, the data demonstrate the important roles of *BCAR4*, through its interaction with SNIP1 and PNUTS, in linking signal-induced acetylation of histone to general transcription machinery during the activation of the GLI2 target genes in breast cancer cells.

BCAR4 as a Potential Therapeutic Target for Breast Cancer Metastasis

To further confirm the functional connection between *BCAR4* and breast cancer metastasis, we performed functional rescue experiments in which we depleted *BCAR4* by LNA, followed by overexpression in MDA-MB-231 cells of either LNA-resistant full-length *BCAR4* or truncated mutants defective for SNIP1 or PNUTS binding (see Figures 2F–2H and S7A). In cell motility assays, knockdown of *BCAR4* reduced migration and invasion of MDA-MB-231 cells, which could be rescued by reintroduction of full-length, but neither $\Delta 212$ – $\Delta 311$ nor $\Delta 968$ – $\Delta 1087$ truncated forms of *BCAR4* (Figures S7B and S7C), even though the expression of full-length *BCAR4* and the truncated forms were equal (Figure S7A), and cell proliferation was not altered (data not shown). Knockdown of *BCAR4* also curtailed the expression of GLI2 target genes and reintroduction of full-length *BCAR4*, but neither $\Delta 212$ – $\Delta 311$ nor $\Delta 968$ – $\Delta 1087$ truncated forms of *BCAR4* were able to robustly rescue the induction of these genes (Figures S7D and S7E). Consistently, knockdown of *BCAR4* abolished CCL21-induced SNIP1 and PNUTS interaction, while reintroduction of full-length *BCAR4*, but neither $\Delta 212$ – $\Delta 311$ nor $\Delta 968$ – $\Delta 1087$ truncated forms of *BCAR4*, was able to robustly rescue the interaction (Figure S7F). These data suggest that *BCAR4* exerts a quantitatively important role in GLI2-dependent

Figure 4. *BCAR4* Is Required for CCL21-Triggered, Phospho-GLI2-Mediated Gene Activation and Cell Migration

(A–C) ChIP quantitative real-time PCR detection of GLI2 (A), phospho-GLI2 (B), or ChIRP quantitative real-time PCR detection of *BCAR4* (C) occupancy on the promoters of selected GLI2 target genes in MDA-MB-231 cells treated with CCL21. RPLP0 served as a non-GLI2 target gene control (A and B). (D) qRT-PCR detection of GLI2 target genes expression in MDA-MB-231 cells transfected with control or *BCAR4* siRNA followed by CCL21 treatment. (E–G) Cell migration assays in MDA-MB-231 cells transfected with indicated siRNA (E and F) or treated with CCL21 neutralization antibody (G). (H) Matrigel cell invasion assay in cells transfected with indicated siRNAs. Left: representative images; right: statistical analysis. (I and J) Matrigel cell invasion assay (I) or qRT-PCR detection of GLI2 target genes (J) in MCF7 cells electroporated with indicated *BCAR4* expression vectors followed by CCL21 treatment.

Error bars, SEM of three independent experiments (* $p < 0.05$, ** $p < 0.01$, and *** $p < 0.001$). See also Figure S4.

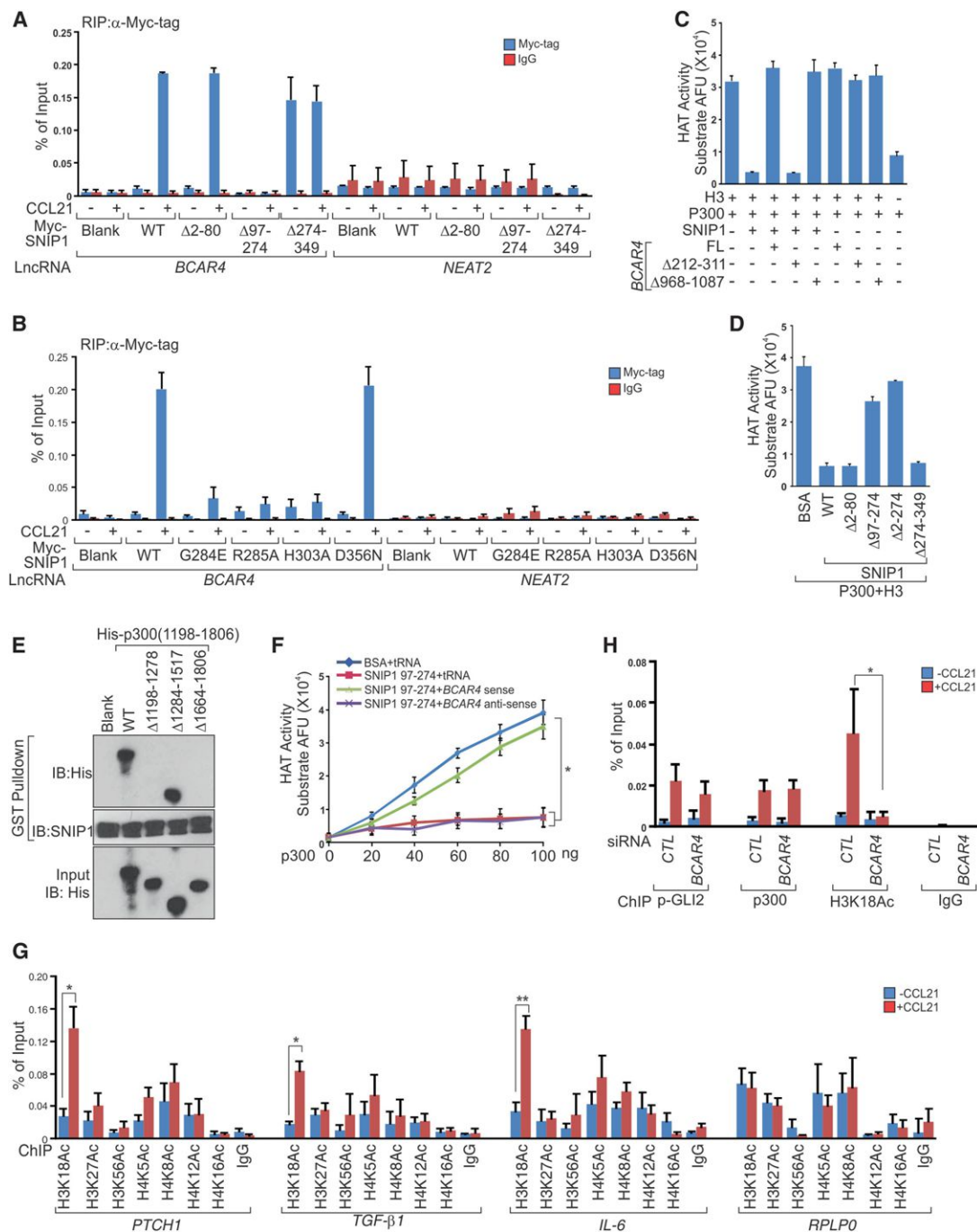


Figure 5. Signal-Induced BCAR4-SNIP1 Interaction Attenuates the Inhibitory Effect of SNIP1 on p300 HAT Activity

(A and B) RIP quantitative real-time PCR detection of the indicated RNAs retrieved by Myc-specific antibody in MDA-MB-231 cells electroporated with indicated vectors followed by CCL21 treatment.

(C) In vitro HAT activity assays of p300 in the presence of WT SNIP1, full-length (FL) BCAR4, and their corresponding mutants as indicated.

(D) In vitro HAT activity assays of p300 in the presence of WT SNIP1 and its corresponding mutants as indicated.

(E) IB detection of the interaction between SNIP1 (aa 97–274) and p300 (aa 1198–1806) WT or truncations.

(F) In vitro HAT activity assays of p300 in the presence of SNIP1 aa 97–274 and BCAR4 sense or antisense RNAs.

(G) ChIP quantitative real-time PCR detection of acetylated histone marks occupancy on the promoters of selected GLI2 target genes in MDA-MB-231 cells treated with CCL21.

(H) ChIP quantitative real-time PCR detection of phospho-GLI2, p300, and H3K18Ac occupancy on *PTCH1* promoter in MDA-MB-231 cells transfected with indicated siRNAs followed by CCL21 treatment.

Error bars, SEM of three independent experiments (* $p < 0.05$ and ** $p < 0.01$). See also Figure S5.

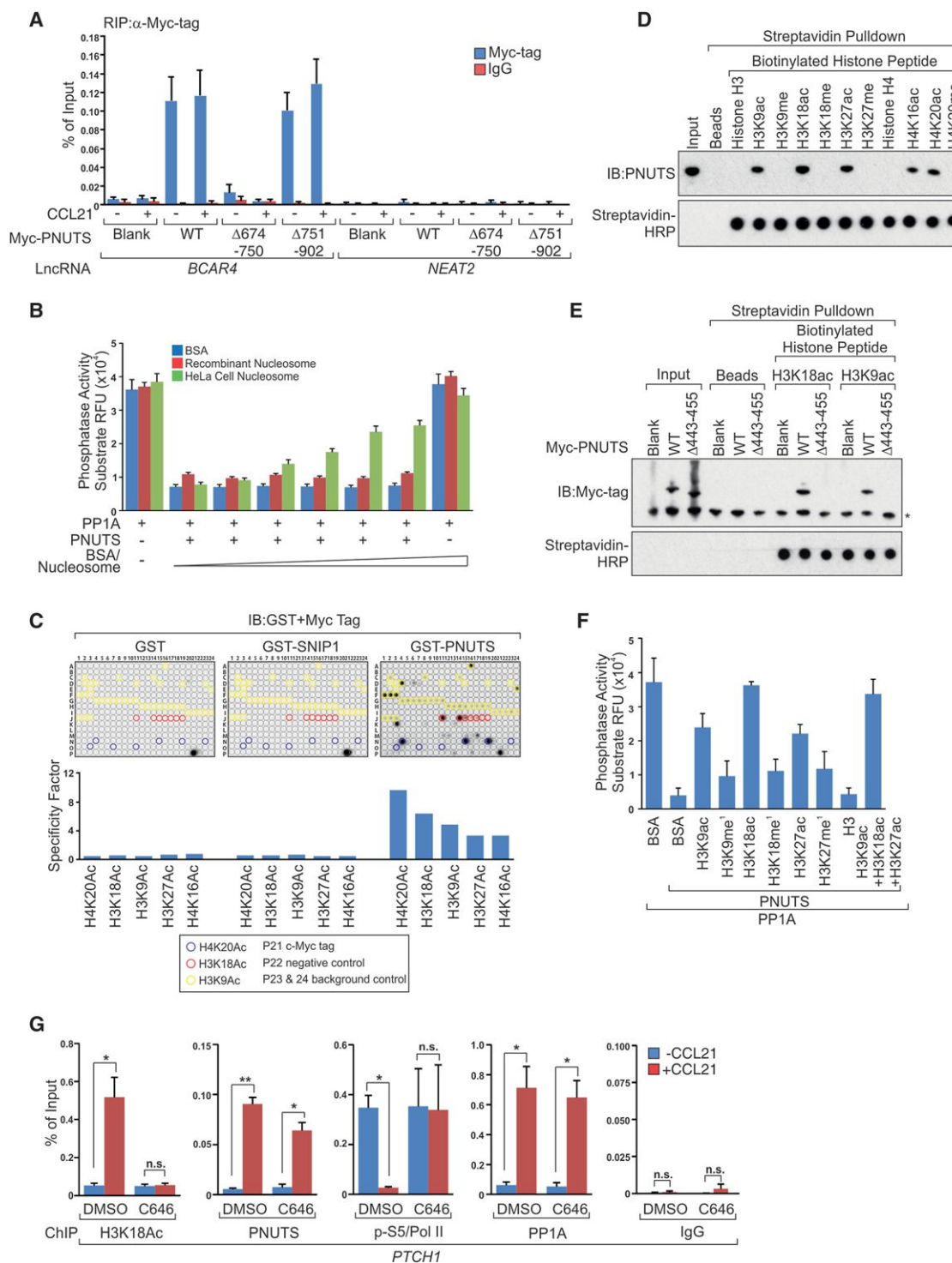


Figure 6. Recognition of *BCAR4*-Dependent Histone Acetylation by PNUTS Attenuates Its Inhibitory Effect on PP1 Activity

(A) RIP quantitative real-time PCR detection of the indicated RNAs retrieved by Myc-specific antibody in MDA-MB-231 cells transfected with indicated vectors followed by CCL21 treatment.

(B) In vitro phosphatase activity assays of PP1A in the presence of indicated proteins or nucleosome.

(C) Modified Histone Peptide Array detection of histone marks recognition by SNIP1 or PNUTS. Top: representative images; bottom: binding specificity.

(D and E) IB detection of PNUTS retrieved by biotinylated histone peptides as indicated from lysate of MDA-MB-231 cells (D) or MDA-MB-231 cells electroporated with indicated vectors (E).

(legend continued on next page)

target gene activation and cell migration/invasion via its direct interactions with SNIP1 and PNUITS.

Next, we recapitulated the contribution of *BCAR4* to breast cancer metastasis in vivo using highly metastatic MDA-MB-231 LM2 cells harboring shRNA-targeting *BCAR4*, which showed reduced migration and invasion (see Figures S4B–S4D). Bioluminescent imaging (BLI) measurements revealed that mammary gland fat-pad injection of MDA-MB-231 LM2 cells harboring control shRNA resulted in lung metastases in NOD/SCID mice, whereas lung metastasis was significantly reduced in two individual groups of mice injected with cells harboring *BCAR4* shRNA (Figure 7A), which was confirmed by quantification of lung metastasis nodules (with an average of 11.2 per mouse in control group and an average of two visible metastases per mouse in *BCAR4* knockdown groups) and histological examination (Figures 7B and 7C). *BCAR4* knockdown had no effect on primary tumor size, tumor cell proliferation, or apoptosis (Figures S7G and S7H), indicating that the metastasis suppression phenotype is not secondary to impaired proliferation or apoptosis. However, CD31, a marker for angiogenesis, was significantly downregulated by *BCAR4* knockdown (Figure S7H), suggesting that reduced lung metastasis burden is due to defective angiogenesis. Independently, the mice with tail-vein injection of *BCAR4* knockdown cells rarely developed lung metastases (Figures 7D–7F). Immunohistochemical analyses confirmed efficient inhibition of metastasis (Figure S7I). These data suggest that *BCAR4* contribute to breast cancer metastasis and silencing of *BCAR4* inhibits lung metastasis in transplantable mouse models.

To evaluate the potential therapeutic potential of *BCAR4*, we synthesized LNAs targeting *BCAR4*. Transfection of LNAs against *BCAR4* into MDA-MB-231 cells exhibited strong knockdown efficiency (see Figure S1I) and dramatically affected cell migration and invasion (data not shown). Next, we examined the therapeutic efficacy of systemically administered in vivo-optimized LNAs in breast cancer metastasis prevention. Of note, two individual LNA treatments significantly reduced lung metastases (Figures 7G and 7H) without notable weight loss (Figure S7J). Importantly, therapeutic LNA-mediated *BCAR4* targeting was confirmed by qRT-PCR analysis of lung metastatic nodules (Figure 7I). Taken together, our findings reveal a *BCAR4*-dependent regulatory network that converges onto a noncanonical hedgehog signaling pathway mediated by phospho-GLI2 to control metastatic initiation and progression in breast cancer.

DISCUSSION

Effective treatment options for breast cancer metastasis, especially for TNBC is not well established. lncRNA-based mechanisms in breast cancer may represent the crucial nodal points for therapeutic intervention. Our studies have revealed that the lncRNA *BCAR4* is highly upregulated in advanced breast cancer patients and contributes to breast cancer metastasis mediated

by chemokine-induced binding of *BCAR4* to two transcription factors with extended regulatory consequences, licensing the activation of a noncanonical hedgehog/GLI2 transcriptional program that promotes cell migration (Figure 7J). In a variety of cancer types, including prostate, breast, ovarian, and pancreatic cancers, hedgehog signaling pathways, which are critical for tumor progression and invasion, are aberrantly activated. We are tempted to speculate that other lncRNAs in these cancer types recognize covalent modifications of GLI2 or other proteins and exert an analogous function to promote the aberrant cancer signaling pathways, which confers cancer cells the invasiveness and metastatic propensity.

Although our data reveal that *BCAR4* exerts a quantitatively important role in chemokine-dependent Hedgehog target gene activation in breast cancer cells, the full mechanisms by which it functions in development remain incompletely defined. *BCAR4* is also highly expressed in human oocyte and placenta (Godinho et al., 2011), suggesting its potential roles in development. Interestingly, hedgehog ligands are expressed in a tissue-specific manner, e.g., desert hedgehog (Dhh) expression is specific to sertoli cells of the testes and granulosa cells of ovaries (Varjosalo and Taipale, 2008). These observations indicate that *BCAR4* is also critical for GLI-mediated gene expression during development.

The *BCAR4* upregulation in breast cancer could be the result of the dysregulation of estrogen receptor (ER). Previous studies have shown that *BCAR4* is upregulated in response to tamoxifen treatment of breast cancer cells (Godinho et al., 2011); thus, upregulation of *BCAR4* could be the result of ER downregulation, as seen in TNBC. It is also possible that *BCAR4* expression is regulated at the transcriptional level by certain aberrant oncogenic signaling pathways in breast cancer cells or by gene amplification at the genomic level. Thus, *BCAR4* expression may require further investigation.

The targeting of lncRNAs with LNAs in breast cancer has not gained much momentum due to the lack of identification of critical breast cancer-relevant lncRNAs and rigorous investigation of the potential anticancer effects of the modulation of lncRNAs in vivo. The important prognostic capacity of *BCAR4* and the robust metastasis suppression by therapeutically delivered LNA targeting *BCAR4* documented in our study encourage future development of lncRNA-based cancer therapies for patients at high risk for metastasis, an outcome currently lacking effective chemotherapeutic options.

EXPERIMENTAL PROCEDURES

lncRNA Array Version 3.0

Total RNA was extracted from two pairs of fresh frozen infiltrating ductal carcinomas of the breast and their adjacent normal breast tissues. RNA samples were subjected to human genome-wide lncRNA microarray (v. 3.0) analyses at ArrayStar. lncRNA array data have been deposited in the Gene Expression Omnibus database under accession GSE60689. Details are included in the Extended Experimental Procedures.

(F) In vitro phosphatase activity assays of PP1A in the presence of PNUITS and modified histones H3 as indicated.

(G) ChIP quantitative real-time PCR detection of H3K18Ac, PNUITS, Pol II Ser5 (normalized by Pol II occupancy), and PP1A occupancy on *PTCH1* promoter in MDA-MB-231 cells pretreated with C646 followed by CCL21 treatment.

Error bars, SEM of three independent experiments (* $p < 0.05$ and ** $p < 0.01$). See also Figure S6.

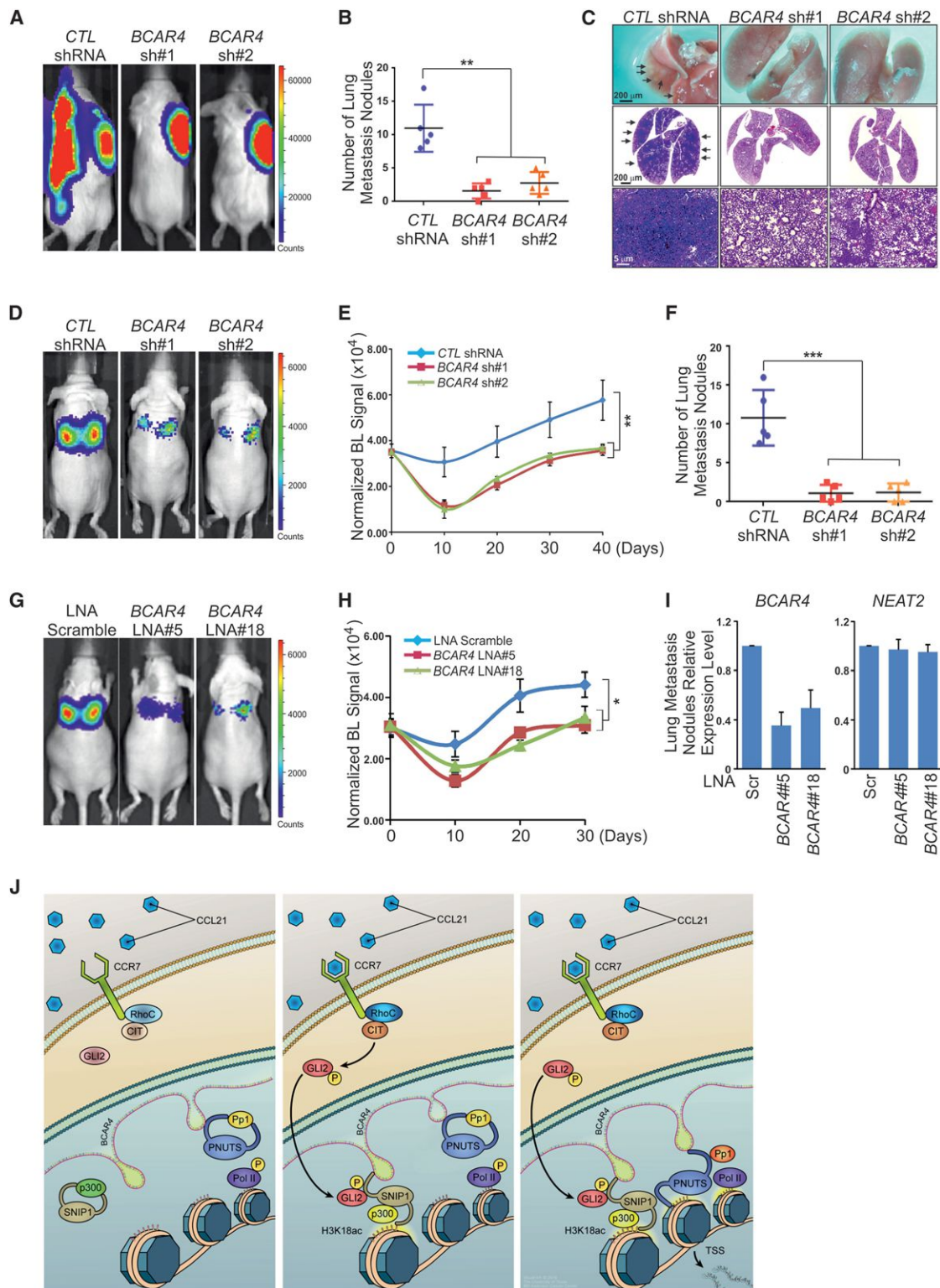


Figure 7. The Potential Therapeutic Role of *BCAR4* in Breast Cancer Metastasis

(A–C) Representative bioluminescent (BL) images (A), metastatic nodules numbers in the lungs (B), or isolated lung bright-field imaging (top) and hematoxylin and eosin staining (middle and bottom) (C) of mice with fat-pad injection of MDA-MB-231 LM2 cells harboring indicated shRNA. Data are means \pm SEM ($n = 5$). (D–F) Representative BLI images (D), lung colonization (E), and metastatic nodules numbers in the lungs (F) of mice with tail-vein injection of MDA-MB-231 LM2 cells harboring indicated shRNA. Data are means \pm SEM ($n = 5$).

(legend continued on next page)

Tissue Specimens

Fresh frozen breast carcinomas and their adjacent normal tissues were purchased from Asterand. Breast cancer tissue microarrays were purchased from Biomax and U.S. BioLab and were grouped into two sets: training set (BC081120, BR1505a, and BR487 from Biomax) and validation set (Bre170Sur-01 from U.S. Biolab). All clinicopathological features of tissue specimens are listed in Table S2. The protocol was approved by the Institutional Review Board at the MD Anderson Cancer Center.

RNA Scope Assay

The RNA Scope probe targeting *BCAR4* was designed and synthesized by Advanced Cell Diagnostics and detection of *BCAR4* expression was performed using the RNA Scope 2.0 High Definition (HD)—BROWN Assay in accordance with the manufacturer's instructions (Advanced Cell Diagnostics). The images were acquired with Zeiss Axioskop2 Plus Microscope.

RNA Pull-Down and Mass Spectrometry Analysis

Biotin-labeled *BCAR4* RNAs were in vitro transcribed with the Biotin RNA Labeling Mix (Roche) and T7 or SP6 RNA polymerase (Ambion) and purified by RNA Clean & Concentrator-5 (Zymo Research). The cell lysates were freshly prepared using ProteaPrep Zwitterionic Cell Lysis Kit, Mass Spec Grade (Protea) with Anti-RNase, Protease/Phosphatase Inhibitor Cocktail, Panobinostat, and Methylstat, supplemented in the lysis buffer. The BcMag Monomer Avidin Magnetic Beads (Bioclone) were first prepared in accordance with manufacturer's instructions and then immediately subjected to RNA (20 μ g) capture in RNA capture buffer (20 mM Tris-HCl [pH 7.5], 1 M NaCl, and 1 mM EDTA) for 30 min at room temperature with agitation. The RNA-captured beads were washed once with NT2 buffer (50 mM Tris-HCl [pH 7.4], 150 mM NaCl, 1 mM $MgCl_2$, and 0.05% NP-40) and incubated with 30 mg cell lysates diluted in NT2 buffer supplemented with 50 U/ml RNase OUT, 50 U/ml Superase•IN, 2 mM dithiothreitol, 30 mM EDTA, and Heparin 0.02 mg/ml for 2 hr at 4°C with rotation. The RNA-binding protein complexes were washed sequentially with NT2 buffer (twice), NT2 high-salt buffer containing 500 mM NaCl (twice), NT2 high-salt buffer containing 1 M NaCl (once), NT2-KSCN buffer containing 750 mM KSCN (twice), and PBS (once) for 5 min at 4°C and eluted by 2 mM D-biotin in PBS. The eluted protein complexes were denatured, reduced, alkylated, and digested with immobilized trypsin (Promega) for MS analysis at MD Anderson Cancer Center Proteomics Facility.

In Vivo Breast Cancer Metastasis Assays

All animal studies were performed with MD Anderson Cancer Center's Institutional Animal Care and Use Committee (IACUC) approval. In vivo spontaneous and experimental breast cancer metastasis assays were performed as described (Chen et al., 2012; Minn et al., 2005). For animal study with LNA injection, mice were intravenously injected with in vivo -grade LNAs (Exiqon) in PBS (15 mg/kg) twice a week for three weeks after MDA-MB-231 LM2 cells injection. The tumor growth and lung metastasis were monitored by Xenogen IVIS 100 Imaging System.

Data Analysis and Statistics

Relative quantities of gene expression level were normalized to *B2M*. The relative quantities of ChIP and ChIRP samples were normalized by individual inputs, respectively. Results are reported as mean \pm SEM of three independent experiments. Comparisons were performed using two-tailed paired Student's *t* test. **p* < 0.05, ***p* < 0.01, and ****p* < 0.001. Fisher's exact test was used for statistical analyses of the correlation between each marker and clinical parameters. For survival analysis, the expression of *BCAR4* was treated as a binary variable, divided into "high" and "low" *BCAR4* expression. Kaplan-Meier sur-

vival curves were compared by the Gehan-Breslow Test in Graphpad Prism (GraphPad Software).

ACCESSION NUMBERS

The Gene Expression Omnibus database accession number for the triple-negative breast cancer LncRNA Array (v. 3.0) reported in this paper is GSE60689.

SUPPLEMENTAL INFORMATION

Supplemental Information includes Extended Experimental Procedures, seven figures, and five tables and can be found with this article online at <http://dx.doi.org/10.1016/j.cell.2014.10.013>.

AUTHOR CONTRIBUTIONS

L.Y. and C. Lin. designed the research, and Z.X., A.L. and C. Li. performed most of the biochemical and molecular experiments. K.L. performed RNA Scope assays and immunohistochemistry analyses. A.L., K.L., and L.Q. performed in vivo tumor xenograft experiments; Y.W. and M.-C.H. performed pathology analyses; and D.H. conducted mass spectrometry analysis. S.W., Y.L., and P.P. assisted in experiments. L.Y. and C. Lin. wrote the manuscript.

ACKNOWLEDGMENTS

We are grateful to Dr. Joan Massague and Dr. Jianming Xu for providing the MDA-MB-231 LM2 cell line and D. Aten for assistance with figure presentation. We thank the Small Animal Imaging Facility at the MD Anderson Cancer Center for assistance in mice imaging. This work was supported by an NIH K99/R00 award (4R00DK094981-02), UT Startup and UT STARS grants (to C. Lin.), and an NIH K99/R00 award (5R00CA166527-03), CPRIT award (R1218), and UT Startup and UT STARS grants (to L.Y.).

Received: February 25, 2014

Revised: July 7, 2014

Accepted: September 24, 2014

Published: November 13, 2014

REFERENCES

- Carpenter, S., Aiello, D., Atianand, M.K., Ricci, E.P., Gandhi, P., Hall, L.L., Byron, M., Monks, B., Henry-Bezy, M., Lawrence, J.B., et al. (2013). A long noncoding RNA mediates both activation and repression of immune response genes. *Science* 341, 789–792.
- Castello, A., Fischer, B., Eichelbaum, K., Horos, R., Beckmann, B.M., Strein, C., Davey, N.E., Humphreys, D.T., Preiss, T., Steinmetz, L.M., et al. (2012). Insights into RNA biology from an atlas of mammalian mRNA-binding proteins. *Cell* 149, 1393–1406.
- Chen, D., Sun, Y., Wei, Y., Zhang, P., Rezaeian, A.H., Teruya-Feldstein, J., Gupta, S., Liang, H., Lin, H.K., Hung, M.C., and Ma, L. (2012). LIFR is a breast cancer metastasis suppressor upstream of the Hippo-YAP pathway and a prognostic marker. *Nat. Med.* 18, 1511–1517.
- Ciurciu, A., Duncalf, L., Jonchere, V., Lansdale, N., Vasieva, O., Glenday, P., Rudenko, A., Vissi, E., Cobbe, N., Alpey, L., and Bennett, D. (2013). PNUTS/PP1 regulates RNAPII-mediated gene expression and is necessary for developmental growth. *PLoS Genet.* 9, e1003885.

(G and H) Representative BLI images (G) and lung colonization (H) of mice at day 30 after tail-vein injection of MDA-MB-231 LM2 cells followed by intravenously LNA treatment. Data are means \pm SEM (*n* = 3).

(I) qRT-PCR detection of *BCAR4* expression in sorted GFP-positive MDA-MB-231 LM2 cells dissociated from lung metastatic nodules of mice intravenously treated with LNA at day 30 (*n* = 3).

(J) A model for cooperative epigenetic regulation downstream of chemokine signals by lncRNA *BCAR4*. See also Figure S7.

- Dias, N., and Stein, C.A. (2002). Antisense oligonucleotides: basic concepts and mechanisms. *Mol. Cancer Ther.* 1, 347–355.
- Dunaeva, M., Michelson, P., Kogerman, P., and Toftgard, R. (2003). Characterization of the physical interaction of Gli proteins with SUFU proteins. *J. Biol. Chem.* 278, 5116–5122.
- Durocher, D., Taylor, I.A., Sarbassova, D., Haire, L.F., Westcott, S.L., Jackson, S.P., Smerdon, S.J., and Yaffe, M.B. (2000). The molecular basis of FHA domain: phosphopeptide binding specificity and implications for phospho-dependent signaling mechanisms. *Mol. Cell* 6, 1169–1182.
- Fatica, A., and Bozzoni, I. (2014). Long non-coding RNAs: new players in cell differentiation and development. *Nat. Rev. Genet.* 15, 7–21.
- Favoni, R.E., and de Cupis, A. (2000). The role of polypeptide growth factors in human carcinomas: new targets for a novel pharmacological approach. *Pharmacol. Rev.* 52, 179–206.
- Feldmann, G., Dhara, S., Fendrich, V., Bedja, D., Beaty, R., Mullendore, M., Karikari, C., Alvarez, H., Iacobuzio-Donahue, C., Jimeno, A., et al. (2007). Blockade of hedgehog signaling inhibits pancreatic cancer invasion and metastases: a new paradigm for combination therapy in solid cancers. *Cancer Res.* 67, 2187–2196.
- Fujii, M., Lyakh, L.A., Bracken, C.P., Fukuoka, J., Hayakawa, M., Tsukiyama, T., Soll, S.J., Harris, M., Rocha, S., Roche, K.C., et al. (2006). SNIP1 is a candidate modifier of the transcriptional activity of c-Myc on E box-dependent target genes. *Mol. Cell* 24, 771–783.
- Gai, M., Camera, P., Dema, A., Bianchi, F., Berto, G., Scarpa, E., Germa, G., and Di Cunto, F. (2011). Citron kinase controls abscission through RhoA and anillin. *Mol. Biol. Cell* 22, 3768–3778.
- Geisler, S., and Collier, J. (2013). RNA in unexpected places: long non-coding RNA functions in diverse cellular contexts. *Nat. Rev. Mol. Cell Biol.* 14, 699–712.
- Godinho, M., Meijer, D., Setyono-Han, B., Dorssers, L.C., and van Agthoven, T. (2011). Characterization of BCAR4, a novel oncogene causing endocrine resistance in human breast cancer cells. *J. Cell. Physiol.* 226, 1741–1749.
- Huarte, M., Guttman, M., Feldser, D., Garber, M., Koziol, M.J., Kenzelmann-Broz, D., Khalil, A.M., Zuk, O., Amit, I., Rabani, M., et al. (2010). A large intergenic noncoding RNA induced by p53 mediates global gene repression in the p53 response. *Cell* 142, 409–419.
- Hui, M., Cazet, A., Nair, R., Watkins, D.N., O'Toole, S.A., and Swarbrick, A. (2013). The Hedgehog signalling pathway in breast development, carcinogenesis and cancer therapy. *Breast Cancer Res.* 15, 203.
- Kakinuma, T., and Hwang, S.T. (2006). Chemokines, chemokine receptors, and cancer metastasis. *J. Leukoc. Biol.* 79, 639–651.
- Kim, R.H., Wang, D., Tsang, M., Martin, J., Huff, C., de Caestecker, M.P., Parks, W.T., Meng, X., Lechleider, R.J., Wang, T., and Roberts, A.B. (2000). A novel smad nuclear interacting protein, SNIP1, suppresses p300-dependent TGF-beta signal transduction. *Genes Dev.* 14, 1605–1616.
- Kim, R.H., Flanders, K.C., Birkey Reffey, S., Anderson, L.A., Duckett, C.S., Perkins, N.D., and Roberts, A.B. (2001). SNIP1 inhibits NF-kappa B signaling by competing for its binding to the C/H1 domain of CBP/p300 transcriptional co-activators. *J. Biol. Chem.* 276, 46297–46304.
- Kim, Y.M., Watanabe, T., Allen, P.B., Kim, Y.M., Lee, S.J., Greengard, P., Naim, A.C., and Kwon, Y.G. (2003). PNU1, a protein phosphatase 1 (PP1) nuclear targeting subunit. Characterization of its PP1- and RNA-binding domains and regulation by phosphorylation. *J. Biol. Chem.* 278, 13819–13828.
- Komarnitsky, P., Cho, E.J., and Buratowski, S. (2000). Different phosphorylated forms of RNA polymerase II and associated mRNA processing factors during transcription. *Genes Dev.* 14, 2452–2460.
- Ling, H., Fabbri, M., and Calin, G.A. (2013). MicroRNAs and other non-coding RNAs as targets for anticancer drug development. *Nat. Rev. Drug Discov.* 12, 847–865.
- Lunde, B.M., Moore, C., and Varani, G. (2007). RNA-binding proteins: modular design for efficient function. *Nat. Rev. Mol. Cell Biol.* 8, 479–490.
- Madaule, P., Eda, M., Watanabe, N., Fujisawa, K., Matsuo, T., Bito, H., Ishizaki, T., and Narumiya, S. (1998). Role of citron kinase as a target of the small GTPase Rho in cytokinesis. *Nature* 394, 491–494.
- Maloverjan, A., Piirsoo, M., Michelson, P., Kogerman, P., and Osterlund, T. (2010). Identification of a novel serine/threonine kinase ULK3 as a positive regulator of Hedgehog pathway. *Exp. Cell Res.* 316, 627–637.
- Meijer, D., van Agthoven, T., Bosma, P.T., Nooter, K., and Dorssers, L.C. (2006). Functional screen for genes responsible for tamoxifen resistance in human breast cancer cells. *Mol. Cancer Res.* 4, 379–386.
- Minn, A.J., Gupta, G.P., Siegel, P.M., Bos, P.D., Shu, W., Giri, D.D., Viale, A., Olshen, A.B., Gerald, W.L., and Massagué, J. (2005). Genes that mediate breast cancer metastasis to lung. *Nature* 436, 518–524.
- Müller, A., Homey, B., Soto, H., Ge, N., Catron, D., Buchanan, M.E., McClanahan, T., Murphy, E., Yuan, W., Wagner, S.N., et al. (2001). Involvement of chemokine receptors in breast cancer metastasis. *Nature* 410, 50–56.
- Rinn, J.L., and Chang, H.Y. (2012). Genome regulation by long noncoding RNAs. *Annu. Rev. Biochem.* 81, 145–166.
- Rubin, L.L., and de Sauvage, F.J. (2006). Targeting the Hedgehog pathway in cancer. *Nat. Rev. Drug Discov.* 5, 1026–1033.
- ten Haaf, A., Bektaş, N., von Serenyi, S., Losen, I., Arweiler, E.C., Hartmann, A., Knüchel, R., and Dahl, E. (2009). Expression of the glioma-associated oncogene homolog (GLI) 1 in human breast cancer is associated with unfavourable overall survival. *BMC Cancer* 9, 298.
- Thélie, A., Papillier, P., Penner, S., Perreau, C., Traverso, J.M., Uzbekova, S., Mermillod, P., Joly, C., Humblot, P., and Dalbiès-Tran, R. (2007). Differential regulation of abundance and deadenylation of maternal transcripts during bovine oocyte maturation in vitro and in vivo. *BMC Dev. Biol.* 7, 125.
- Varjosalo, M., and Taipale, J. (2008). Hedgehog: functions and mechanisms. *Genes Dev.* 22, 2454–2472.
- Wang, Z., Zang, C., Rosenfeld, J.A., Schones, D.E., Barski, A., Cuddapah, S., Cui, K., Roh, T.Y., Peng, W., Zhang, M.Q., and Zhao, K. (2008). Combinatorial patterns of histone acetylations and methylations in the human genome. *Nat. Genet.* 40, 897–903.
- Wang, X., Venugopal, C., Manoranjan, B., McFarlane, N., O'Farrell, E., Nolte, S., Gunnarsson, T., Hollenberg, R., Kwiecien, J., Northcott, P., et al. (2012). Sonic hedgehog regulates Bmi1 in human medulloblastoma brain tumor-initiating cells. *Oncogene* 31, 187–199.
- Washington, K., Ammosova, T., Beullens, M., Jerebtsova, M., Kumar, A., Bolten, M., and Nekhai, S. (2002). Protein phosphatase-1 dephosphorylates the C-terminal domain of RNA polymerase-II. *J. Biol. Chem.* 277, 40442–40448.
- Yamashiro, S., Totsukawa, G., Yamakita, Y., Sasaki, Y., Madaule, P., Ishizaki, T., Narumiya, S., and Matsumura, F. (2003). Citron kinase, a Rho-dependent kinase, induces di-phosphorylation of regulatory light chain of myosin II. *Mol. Biol. Cell* 14, 1745–1756.
- Yang, L., Lin, C., Jin, C., Yang, J.C., Tanasa, B., Li, W., Merkurjev, D., Ohgi, K.A., Meng, D., Zhang, J., et al. (2013). lncRNA-dependent mechanisms of androgen-receptor-regulated gene activation programs. *Nature* 500, 598–602.
- Yu, B., Bi, L., Zheng, B., Ji, L., Chevalier, D., Agarwal, M., Ramachandran, V., Li, W., Lagrange, T., Walker, J.C., and Chen, X. (2008). The FHA domain proteins DAWDLE in Arabidopsis and SNIP1 in humans act in small RNA biogenesis. *Proc. Natl. Acad. Sci. USA* 105, 10073–10078.

CDK7 Inhibition Suppresses Super-Enhancer-Linked Oncogenic Transcription in MYCN-Driven Cancer

Edmond Chipumuro,^{1,2} Eugenio Marco,^{3,4} Camilla L. Christensen,⁵ Nicholas Kwiatkowski,⁶ Tinghu Zhang,^{7,8} Clark M. Hatheway,¹ Brian J. Abraham,⁶ Bandana Sharma,¹ Caleb Yeung,^{1,2} Abigail Altabel,⁵ Antonio Perez-Atayde,⁹ Kwok-Kin Wong,⁵ Guo-Cheng Yuan,^{3,4} Nathanael S. Gray,^{7,8} Richard A. Young,⁶ and Rani E. George^{1,2,*}

¹Department of Pediatric Hematology/Oncology, Dana-Farber Cancer Institute and Boston Children's Hospital, Boston, MA 02215, USA

²Department of Pediatrics, Harvard Medical School, Boston, MA 02115, USA

³Department of Biostatistics and Computational Biology, Dana-Farber Cancer Institute, Boston, MA 02215, USA

⁴Harvard School of Public Health, Boston, MA 02115, USA

⁵Department of Medical Oncology, Dana-Farber Cancer Institute, Boston, MA 02215, USA

⁶Whitehead Institute for Biomedical Research and Department of Biology, Massachusetts Institute of Technology, Cambridge, MA 02139, USA

⁷Department of Cancer Biology, Dana-Farber Cancer Institute, Boston, MA 02215, USA

⁸Department of Biological Chemistry and Molecular Pharmacology, Harvard Medical School, Boston, MA 02115, USA

⁹Department of Pathology, Boston Children's Hospital, MA 02115, USA

*Correspondence: rani_george@dfci.harvard.edu

<http://dx.doi.org/10.1016/j.cell.2014.10.024>

SUMMARY

The MYC oncoproteins are thought to stimulate tumor cell growth and proliferation through amplification of gene transcription, a mechanism that has thwarted most efforts to inhibit MYC function as potential cancer therapy. Using a covalent inhibitor of cyclin-dependent kinase 7 (CDK7) to disrupt the transcription of amplified *MYCN* in neuroblastoma cells, we demonstrate downregulation of the oncoprotein with consequent massive suppression of MYCN-driven global transcriptional amplification. This response translated to significant tumor regression in a mouse model of high-risk neuroblastoma, without the introduction of systemic toxicity. The striking treatment selectivity of *MYCN*-overexpressing cells correlated with preferential downregulation of super-enhancer-associated genes, including *MYCN* and other known oncogenic drivers in neuroblastoma. These results indicate that CDK7 inhibition, by selectively targeting the mechanisms that promote global transcriptional amplification in tumor cells, may be useful therapy for cancers that are driven by MYC family oncoproteins.

INTRODUCTION

Many human cancers depend on the deregulated expression of MYC family members for their aberrant growth and proliferation, with elevated expression of these oncogenes predicting aggressive disease and a poor clinical outcome (Eilers and Eisenman, 2008; Wasylishen and Penn, 2010). Deactivation of MYC in cell lines and MYC-induced transgenic tumors causes proliferative

arrest and tumor regression (Arvanitis and Felsher, 2006; Soucek et al., 2008), suggesting that effective targeting of MYC proteins would have broad therapeutic benefit. Recently, several groups reported that oncogenic MYC elicits its plethora of downstream effects in tumor cells through global transcriptional amplification, leading to massively upregulated expression of genes involved in multiple processes (Lin et al., 2012; Lovén et al., 2012; Nie et al., 2012; Schuhmacher and Eick, 2013). When present at physiological levels, MYC binds to the core promoters of actively transcribed genes; however, in tumor cells with MYC overexpression, increased MYC levels are observed at both the core promoters and enhancers of the same set of genes, resulting in increased levels of transcripts per cell. This mechanism provides an explanation for the lack of a common MYC transcriptional signature and for the diverse effects of deregulated MYC in cancer cells. Another general feature of deregulated MYC is its transcriptional regulation by super-enhancers (SEs), clusters of enhancers that are densely occupied by transcription factors, cofactors, and chromatin regulators (Hnisz et al., 2013). SEs are acquired by cancer cells through gene amplification, translocation or transcription factor overexpression. They facilitate high-level expression of genes, including MYC, whose protein products are critical for the control of cell identity, growth, and proliferation, and which are especially sensitive to perturbation (Chapuy et al., 2013; Hnisz et al., 2013; Lovén et al., 2013). These emerging insights into the role of oncogenic MYC as an SE-associated transcriptional amplifier suggest that strategies aimed at disrupting the molecular mechanisms that drive this function might provide useful therapy for different MYC-dependent tumors.

The transcription cycle of RNA polymerase II (Pol II) is regulated by a set of cyclin-dependent kinases (CDKs) that have critical roles in transcription initiation and elongation (Larochelle et al., 2012). In contrast to the cell-cycle CDKs which are largely responsible for cell-cycle transition, these transcriptional CDKs

(especially CDK7, a subunit of TFIIF, and CDK9, a subunit of pTEFb) phosphorylate the carboxy-terminal domain (CTD) of Pol II, facilitating efficient transcriptional initiation, pause release and elongation. Moreover, most CDKs are activated through T-loop phosphorylation by a CDK-activating kinase (CAK), which in metazoans appears to be uniquely controlled by CDK7 (Fisher and Morgan, 1994; Glover-Cutter et al., 2009; Larochelle et al., 2007, 2012; Rossignol et al., 1997; Serizawa et al., 1995). Inhibition of transcriptional CDKs primarily affects the accumulation of transcripts with short half-lives, including antiapoptosis family members and cell-cycle regulators (Garriga and Graña, 2004; Lam et al., 2001), rendering this group of kinases ideal candidates for blocking MYC-dependent transcriptional amplification.

Here, we investigate whether inhibition of transcriptional CDKs can be exploited to disrupt aberrant MYC-driven transcription, using the deregulated expression of *MYCN* as a model. The *MYCN* protein shares most of the physical properties of MYC (Kohl et al., 1986) and is considered functionally interchangeable, based on the similarity of their transcriptional programs, the cellular phenotypes they induce, and the ability of *MYCN* to replace MYC during murine development (Boon et al., 2001; Malynn et al., 2000; Toyoshima et al., 2012). In neuroblastoma (NB), a pediatric solid tumor arising in the peripheral sympathetic nervous system, *MYCN* amplification is typically associated with a dismal prognosis, regardless of the treatment used (Brodeur et al., 1984; Seeger et al., 1985).

We demonstrate that THZ1, a newly developed covalent inhibitor of CDK7 (Kwiatkowski et al., 2014), selectively targets *MYCN*-amplified NB cells, leading to global repression of *MYCN*-dependent transcriptional amplification. This response induces sustained growth inhibition of tumors in a mouse model of NB. The remarkable sensitivity of *MYCN*-amplified cells to CDK7 inhibition is associated with preferentially reduced expression of SE-associated oncogenic drivers, especially *MYCN*.

RESULTS

THZ1 Exerts Potent and Selective Effects in *MYC*-Deregulated Cells

To identify CDKs whose depletion might lead to decreased *MYCN* expression and subsequent apoptosis, we performed a short hairpin (shRNA) knockdown screen of CDKs with known transcriptional activities (CDK7, CDK8, CDK9, CDK12, CDK13, and CDK19) in Kelly cells, a human NB line that expresses very high levels of *MYCN* RNA and protein due to genomically amplified *MYCN* (100–120 copies per cell) (Schwab et al., 1983). Genetic depletion of CDK7, CDK8, CDK9, or CDK19 led to marked decreases of *MYCN* RNA and protein with a concomitant increase in cleaved caspase 3 (CC3) expression (Figures S1A and S1B available online). To reproduce these results pharmacologically, we tested a panel of 11 inhibitors with activity against transcriptional CDKs in three *MYCN*-amplified cell lines, observing a range of sensitivities, with the highest potency (IC_{50} , 6–9 nM) shown by a newly developed covalent phenylaminopyrimidine inhibitor of CDK7, THZ1 (Kwiatkowski et al., 2014) (Figure S1C; Table S1). Similar results were obtained when THZ1 was tested against a larger panel of *MYCN*-amplified NB cell lines with varying levels of *MYCN* expression (Figures 1A and

S1D). Importantly, NB cells without *MYCN* amplification were far less sensitive to THZ1, with IC_{50} values averaging ten times higher than those of *MYCN*-amplified cells (Figure 1A; Table S1). Notably, the NBL-S cell line, which expresses high levels of *MYCN* without genomic amplification (Cohn et al., 1990) (Figure S1D), was quite sensitive to THZ1, while two nontransformed lines (B6-MEFs and NIH 3T3) were relatively insensitive (Figure 1A). *MYCN*-amplified cells also showed enhanced sensitivity to THZ1R, a reversible analog of THZ1 that lacks the acrylamide moiety required for covalent bond formation, although it was not as potent as the covalent inhibitor (Figure 1B). The strong selectivity of THZ1 for *MYCN*-amplified cells was not restricted to NB, but extended to H262-BT111, a human primitive neuroectodermal tumor cell line expressing amplified *MYCN* (K. Ligon, personal communication) and Raji and Daudi lymphoma cells, both characterized by *MYC* overexpression due to chromosomal translocation (Nishikura et al., 1985; Veronese et al., 1995) (Figure S1E). To probe this preferential effect further, we tested the extent of target engagement in NB cells, using a biotinylated derivative of THZ1 (bio-THZ1) with or without THZ1 pretreatment. Bio-THZ1 consistently bound to CDK7 in both *MYCN*-amplified and nonamplified untreated cells (Figure S1F), but became less efficient after THZ1 treatment, suggesting that target recognition was not a major factor in the enhanced inhibitory effects of THZ1 in *MYCN*-amplified cells.

Next, we studied the growth inhibitory effects of THZ1. *MYCN*-amplified cells treated with THZ1 underwent cell-cycle arrest in G2/M at 24 hr, an effect that was not observed in *MYCN*-nonamplified NB cells, even after 48 hr (Figure 1C). Moreover, THZ1 led to a profound induction of apoptosis in high *MYCN*-expressing cells, but not in cells expressing nonamplified *MYCN* (Figures 1D and S1G). Together, these data indicate that THZ1 induces selective cytotoxicity not only in NB cells with *MYCN* amplification, but also in other cancers overexpressing either the *MYCN* or *MYC* oncogene.

THZ1 Inhibits Tumor Growth in a Mouse Model of Human *MYCN*-Amplified NB

Given the relative lack of target specificity of past CDK inhibitors, leading to adverse effects in normal cells (Lapenna and Giordano, 2009), we assessed the tolerability of THZ1 in non-tumor-bearing mice ($n = 6$) treated with 10 mg/kg intravenously twice daily. No systemic toxicity was observed even after 4 weeks of continuous administration (data not shown). We next tested the therapeutic effects of THZ1 in xenograft models of *MYCN*-amplified human NB derived from subcutaneous flank injection of Kelly cells. When tumors reached an optimal size (mean volume, ~ 150 mm³; range, 75–235 mm³), the animals were divided into two groups and treated with vehicle ($n = 9$) or THZ1 as above ($n = 14$). Treatment was continued for a mean of 20 days (range, 15–24 days) in the control group and 24 days (range, 20–28 days) in the THZ1 group. Mice receiving THZ1 had a statistically significant reduction in tumor growth, again without toxicity (Figure 1E). Two animals remained free of tumor recurrence at 35 and 128 days posttreatment.

Tumors from vehicle-treated mice displayed histological features of human NB with poorly differentiated, small round blue cells displaying high mitotic activity (Figure 1F). By contrast,

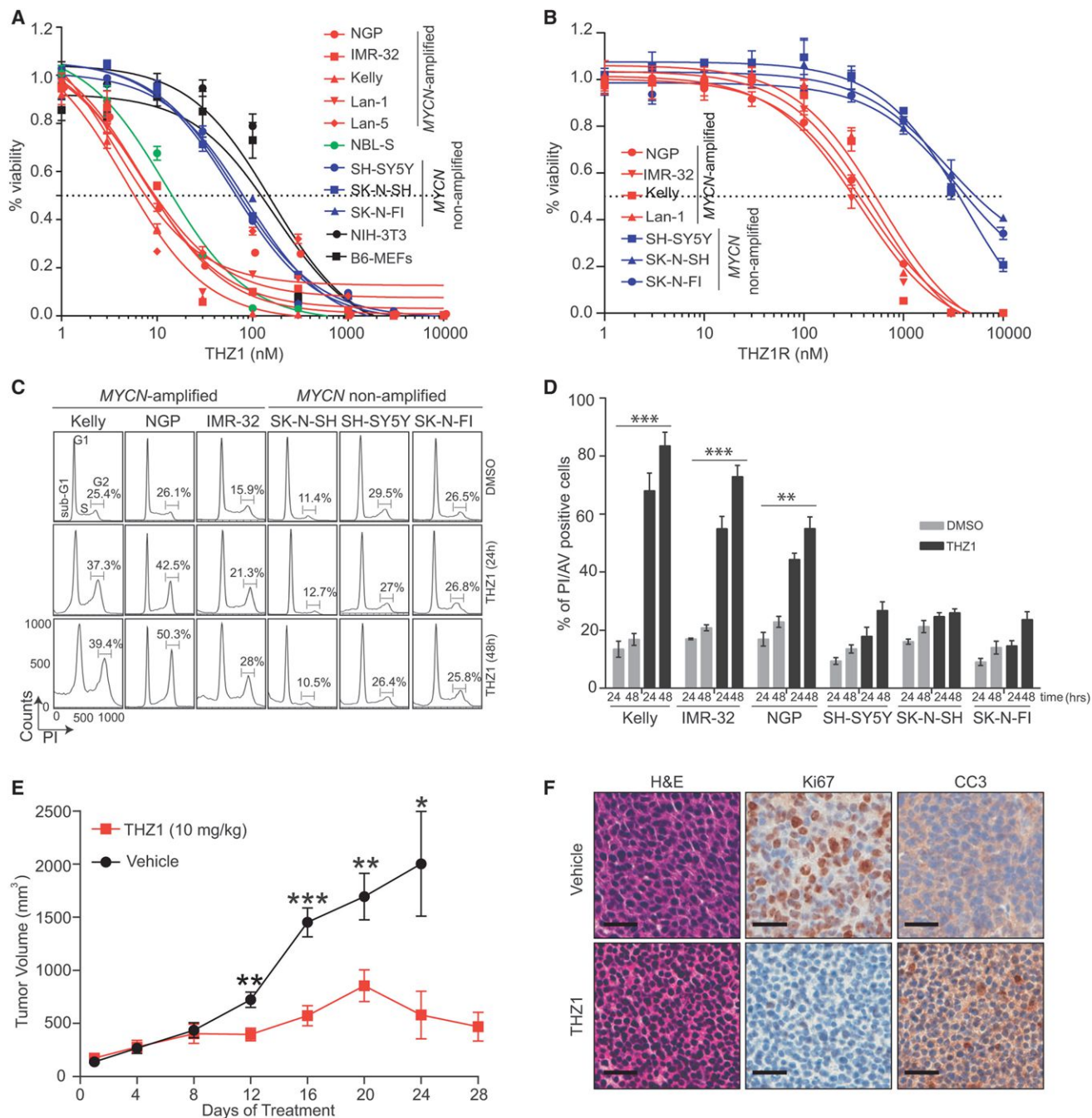


Figure 1. THZ1 Exhibits High Potency and Selectivity against MYCN-Amplified Tumor Cells

(A) Dose-response curves of MYCN-amplified and nonamplified human NB and murine fibroblast cells after treatment with increasing concentrations of THZ1 for 72 hr. Percent cell viability relative to that of DMSO-treated cells is shown here and in (B). Data represent mean \pm SD of three replicates here and in (B).

(B) Dose-response curves of NB cells treated as in (A) with the reversible CDK7 inhibitor THZ1R.

(C) Cell-cycle analysis of MYCN-amplified versus nonamplified NB cells exposed to THZ1 (100 nM \times 24 and 48 hr) by flow cytometry with propidium iodide (PI) staining. Results are representative of three replicates. The scale and axes are indicated in the lower left corner.

(D) Apoptosis analysis in MYCN-amplified and nonamplified NB cells treated with THZ1 as in (C) by flow cytometry with Annexin V staining. Data represent mean \pm SD of three replicates. *** p < 0.0001, ** p < 0.001 (Student's t test).

(E) Tumor volumes of MYCN-amplified human NB xenografts in NU/NU (*Crl:NU-Foxn1^{nu}*) mice treated with THZ1 (10 mg/kg intravenously [i.v.] twice daily) (n = 14) or vehicle (n = 9) for 28 days. Mean \pm SD values are presented. *** p < 0.001; ** p < 0.01; * p < 0.05 (multiple t test, Holm-Sidak method).

(F) Immunohistochemical (IHC) analysis of morphology (hematoxylin & eosin [H&E]), proliferation (Ki67) and apoptosis (cleaved caspase 3 [CC3]) in tumors harvested from animals treated with vehicle or THZ1 as in (E) for 12 days. Scale bar represents 25 μ m.

See also Figure S1 and Table S1.

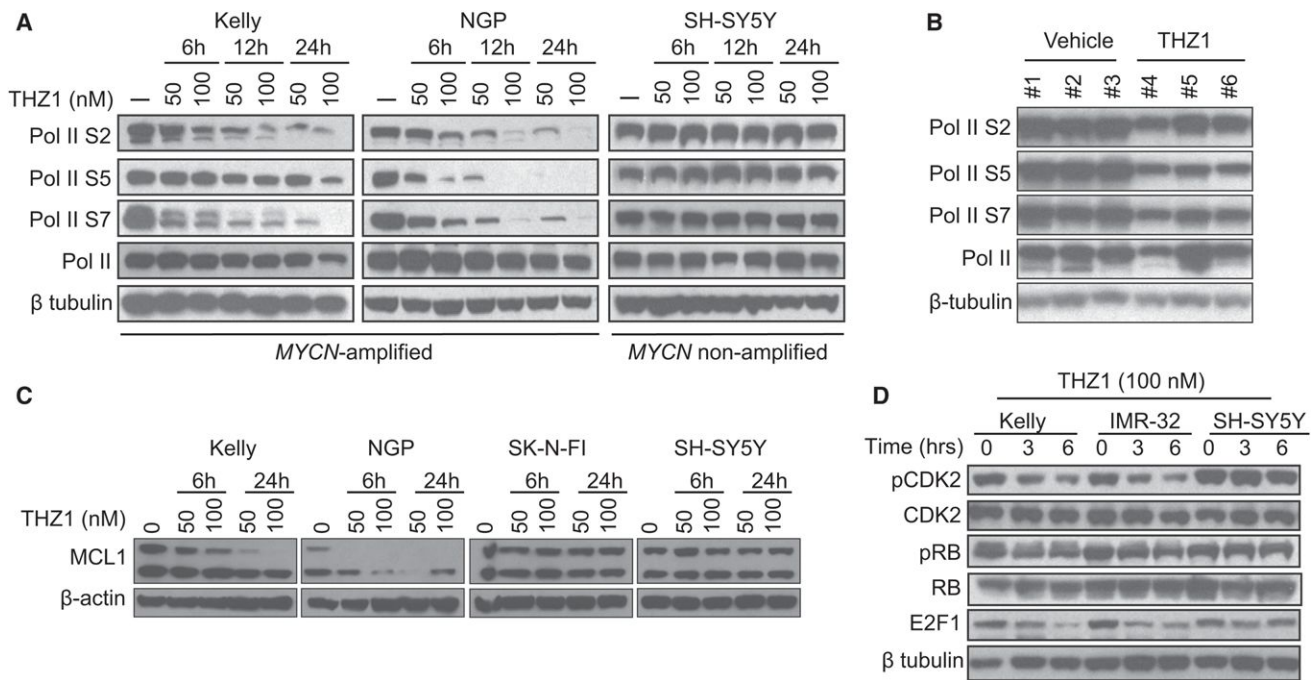


Figure 2. THZ1 Inhibits General Transcription and Cell-Cycle Regulation in MYCN-Amplified Tumor Models

(A) Immunoblot analysis of RNA Pol II CTD phosphorylation in MYCN-amplified and nonamplified NB cells treated with DMSO or THZ1 at the indicated concentrations for the indicated times.

(B) Immunoblot analysis of RNA Pol II CTD phosphorylation in human NB xenograft tumor cells obtained from mice treated with vehicle or THZ1 (10 mg/kg i.v. twice daily) for 12 days.

(C) Immunoblot analysis of MCL1 in MYCN-amplified (Kelly, NGP) versus nonamplified (SK-N-FI, SH-SY5Y) NB cells following treatment with THZ1 at the indicated concentrations and durations.

(D) Immunoblot analysis of proteins involved in cell-cycle progression in MYCN-amplified (Kelly, IMR-32) and nonamplified (SH-SY5Y) NB cells following treatment with THZ1 100 nM for 3 and 6 hr.

See also Figure S2.

the vast majority of tumor cells in the THZ1-treated animals demonstrated necrosis, reduced proliferative activity and increased apoptosis. To ensure target engagement in the tumor cells, we used bio-THZ1 to pull down CDK7 in cell lysates from both vehicle- and THZ1-treated animals, noting decreased binding with bio-THZ1 in the latter (Figure S1H). We also confirmed that the lack of toxicity in the animal models did not reflect THZ1 selectivity for human CDK7, as bio-THZ1 formed a complex with murine CDK7 in cell lysates from NIH 3T3 cells treated with THZ1 (Figure S1H). These results demonstrate the feasibility of specifically targeting CDK7 in tumor cells driven by deregulated *MYC* or *MYCN*.

THZ1 Affects Both Transcription and Cell-Cycle Progression in MYCN-Amplified Cells

As a transcriptional kinase, CDK7 exerts its effects through regulation of RNA Pol II-mediated transcriptional initiation and pause establishment, while also affecting elongation through its CAK activity on other transcriptional CDKs (Glover-Cutter et al., 2009; Larochelle et al., 2012; Palancade and Bensaude, 2003). We observed a dose-dependent decrease in the initiation-associated serine 5 (S5) and serine 7 (S7) and the elongation-associated serine 2 (S2) Pol II phosphorylation in MYCN-amplified but not nonamplified cells treated with THZ1 (Figures 2A and S2A).

Pol II CTD phosphorylation was also downregulated in tumor cells from animals treated with THZ1 (Figure 2B). Downregulation of CDK7 phosphorylation was seen in MYCN-amplified cells (Figure S2B), consistent with the finding that CDK7 is regulated by phosphorylation within its own activation (T) loop (Larochelle et al., 2012). Phosphorylation of CDK9 was also decreased in MYCN-amplified cells (Figure S2C), reinforcing the effect of THZ1 on transcription elongation. Total protein levels of CDK9 were also decreased in these cells but not nonamplified cells, suggesting that THZ1-induced CDK7 inhibition might also target the transcription of CDK9 (Figure S2C).

The decrease in Pol II phosphorylation after THZ1 treatment in MYCN-amplified cells coincided with the loss of the short-lived antiapoptotic protein MCL1 (Figure 2C). Similar effects on Pol II CTD phosphorylation and MCL1 levels were observed in Raji, Daudi, and H262-BTIII cells, all of which depend on *MYC* or *MYCN* overexpression and are sensitive to THZ1 (Figure S2D). Importantly, the effects of THZ1, including Pol II CTD phosphorylation and the induction of apoptosis, could be rescued by a mutation in the covalent binding site of CDK7 (*CDK7^{C312S}*) (Figure S2E), indicating on-target effects of THZ1 on CDK7.

CDK7 also stimulates cell-cycle progression by activating CDK1 and CDK2 through its T-loop phosphorylation function (Larochelle et al., 2007). We observed a time-dependent

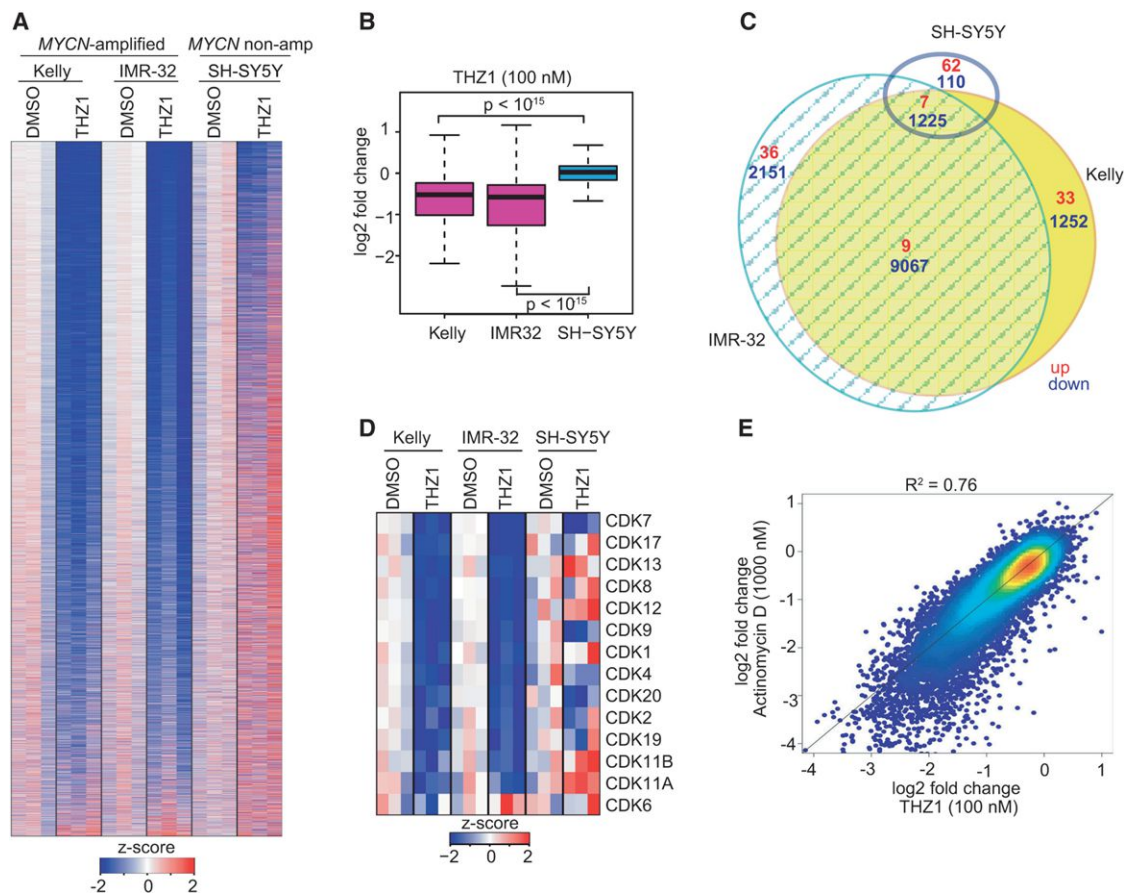


Figure 3. THZ1 Causes Massive Downregulation of Actively Transcribed Genes in MYCN-Overexpressing NB Cells

(A) Heatmap of gene expression values in MYCN-amplified and nonamplified cells treated with THZ1 (100 nM for 6 hr) versus DMSO. Rows show Z scores calculated for each cell type.

(B) Quartile box plots of log2 fold changes in gene expression in MYCN-amplified and nonamplified cells treated with DMSO or THZ1 at the same dose and duration as in (A). Box plot whiskers extend to 1.5 times the interquartile range (n = 18,665 expressed genes, p < 10⁻¹⁵ for Kelly versus SH-SY5Y and IMR-32 versus SH-SY5Y, two-sided Mann-Whitney U test).

(C) Venn diagram depicting the overlap between sets of differentially expressed transcripts (THZ1 versus DMSO) in MYCN-amplified (Kelly, IMR-32), and nonamplified (SH-SY5Y) cells treated with THZ1 as in (A). Red represents upregulated transcripts. Blue represents downregulated transcripts.

(D) Heatmap of gene expression values of transcriptional and cell-cycle CDKs in MYCN-amplified and nonamplified cells treated with THZ1 as in (A) versus DMSO. Rows show Z scores calculated for each cell type.

(E) Correlation between log2 fold changes in gene expression following THZ1 (100 nM) versus DMSO treatment and actinomycin D (1 μM) versus DMSO treatment for 6 hr. in MYCN-amplified NB cells. R² (coefficient of determination) calculated using a simple linear regression model.

See also Figure S3.

decrease in CDK1 and CDK2 phosphorylation and in other proteins involved in cell-cycle regulation (pRB and E2F), that was more striking in MYCN-amplified cells (Figures 2D and S2F). However, the Pol II inhibitor triptolide appeared to show selectivity for MYCN-overexpressing cells, compared to purvanolol, which primarily targets cell-cycle CDKs (Figure S2G). Thus, although THZ1 causes cytotoxicity through transcriptional as well as cell-cycle inhibition in MYC-driven tumor cells, the latter effect is most likely through decreased transcription of cell-cycle CDKs.

CDK7 Inhibition Causes Widespread Transcriptional Shutdown in MYCN-Amplified NB Cells

Because THZ1 preferentially downregulates RNA Pol II CTD phosphorylation in MYCN-amplified cells (Figures 2A and S2A),

we next compared the consequences of CDK7 inhibition on global gene expression in MYCN-amplified (Kelly and IMR-32) versus nonamplified (SH-SY5Y) cells following exposure to 100 nM THZ1 for 6 hr. Expression profiling was performed with “spike-in” RNA standards normalized to cell number to enable accurate detection of differences in total RNA levels in cells with or without MYCN overexpression (Lovén et al., 2012).

We observed a genome-wide decrease in the mRNA expression levels of most actively transcribed genes in MYCN-amplified cell lines compared to DMSO-treated cells (Figure 3A). This widespread downregulation of gene expression was not apparent in MYCN-nonamplified SH-SY5Y cells treated with the same dose and duration of THZ1 (Figure 3A). Indeed, the percentage of actively transcribed genes significantly downregulated in

these cells was only 7.5%, contrasting with 63% for Kelly and 68% for IMR-32 *MYCN*-amplified cells (false discovery rate [FDR] <0.05) (Figure 3B). Moreover, Kelly and IMR-32 cells had 55% ($n = 10,292$) downregulated transcripts in common, while only 12% ($n = 1,225$) of these were also downregulated in SH-SY5Y cells (out of 18,665 genes) (Figure 3C). The most downregulated genes and gene sets in the *MYCN*-amplified cells were transcriptional and cell-cycle regulators (Figures S3A and S3B); including transcripts of CDKs and their partner cyclins (Figures 3D and S3C). Comparison of expression profiles of *MYCN*-amplified cells exposed to THZ1 with those exposed to the general transcription inhibitor actinomycin D revealed a significant correlation between the two ($R^2 = 0.76$) (Figure 3E). These results led us to conclude that CDK7 inhibition, through THZ1, preferentially downregulates global transcription in *MYCN*-amplified NB cells, affecting most active genes.

THZ1 Targets Deregulated *MYCN* in NB Cells

The selectivity of THZ1 for *MYC/MYCN*-deregulated cells led us to investigate its effects on *MYCN* expression and function. We observed that *MYCN* mRNA was among the top 15% of significantly downregulated transcripts in NB cell lines most sensitive to THZ1. Moreover, transcripts differentially regulated on exposure to THZ1 showed enrichment for previously published *MYCN/MYC* target gene sets (Figure S4A). THZ1 treatment resulted in a dose- and time-dependent downregulation of *MYCN* mRNA and protein levels in *MYCN*-amplified cells (Figures 4A and 4B), which were rescued by overexpression of the *CDK7*^{C312S} mutant (Figure S4B), similar to observations with an ATP analog in colorectal carcinoma cells expressing a *CDK7* mutation (Glover-Cutter et al., 2009). We also observed downregulation of *MYCN* protein levels in the human NB tumors from mice that responded to THZ1 (Figure 4C). These effects primarily reflected a reduction in *MYCN* transcripts, as comparison of *MYCN* protein turnover in *MYCN*-amplified cells exposed to either THZ1 or actinomycin D showed largely similar rates of degradation (Figure S4C). Finally, analysis of *MYCN* occupancy at the promoters of two of its known transcriptional targets, *MDM2* (Slack et al., 2005) and *MCL1* (Labisso et al., 2012), revealed significantly decreased *MYCN* binding, consistent with the disruption of *MYCN* function by THZ1 (Figure 4D).

Next, we determined whether genetic depletion of *MYCN* mimicked the effects of THZ1 in *MYCN*-amplified cells. Abrogation of amplified *MYCN* expression in Kelly cells using shRNA knockdown led to decreased Pol II CTD phosphorylation at S2, S5, and S7 (Figure 4E), similar to our findings with THZ1 treatment. The inhibition of transcription associated with *MYCN* knockdown was coupled with complete loss of *MCL1* and subsequent cell death (Figure 4E). We next asked whether ectopic expression of *MYCN* would sensitize cells to THZ1. *MYCN*-nonamplified SH-SY5Y cells were engineered to express the oncogene by lentiviral transduction, resulting in approximately 380-fold increases in *MYCN* mRNA and protein levels (Figure 4F), equivalent to those in *MYCN*-amplified cells (Figure S1D). Ectopic expression of *MYCN* in these cells led to a 5-fold greater sensitivity to THZ1 compared with untransfected or vector control-transfected SH-SY5Y cells (Figure 4G). Together, these

results indicate that THZ1 specifically targets deregulated *MYCN* in NB cells.

Amplified *MYCN* Induces Global Transcriptional Amplification that Is Inhibited by THZ1

Downregulation of *MYCN* alone seemed insufficient to account for the widespread effects of THZ1 on the transcriptional output of cells harboring amplified *MYCN*. We therefore questioned whether deregulated *MYCN*, like *MYC*, binds promiscuously to active genes, leading to global transcriptional amplification, and whether THZ1 preferentially disrupts this effect. Indeed, ectopic expression of *MYCN* in SH-SY5Y cells led to increased cell size (Figure S5A) and significantly increased expression of all active transcripts (Figures 5A and S5B). Moreover, THZ1 treatment led to significant downregulation of transcriptional output in these cells (Figure 5B) with induction of apoptosis, although *MYCN* protein levels were not affected (Figure 5C). Furthermore, shRNA knockdown of *MYCN* in *MYCN*-amplified cells led to a striking downregulation of global gene expression in comparison to their control shRNA-expressing counterparts (Figures 5D and 5E), again underscoring the widespread transcriptional effects of deregulated *MYCN*. To determine the extent to which these effects mirrored those of CDK7 inhibition, we compared the expression profiles of *MYCN*-amplified cells expressing a *MYCN* shRNA with those of THZ1-treated cells, noting significant downregulation of the actively transcribed genes in both types of cells, with a close correlation between the top 500 downregulated transcripts in THZ1-treated cells and those in *MYCN* shRNA-depleted cells (Figures 5F and S5C). Hence, disruption of *MYCN*-induced transcriptional amplification also contributes to the striking effects of THZ1.

Super-Enhancers Account for the Selectivity of THZ1 for *MYCN*-Amplified NB Cells

We observed that the ectopic overexpression of *MYCN* in nonamplified cells led to global transcriptional upregulation and sensitization to the growth inhibitory effects of THZ1, although the latter effect did not approach the levels seen in cells with endogenously amplified *MYCN*. We therefore investigated the possible contribution of super-enhancers (SEs) to the extreme susceptibility of endogenous *MYCN*-amplified cells to THZ1. Using chromatin immunoprecipitation with high-throughput sequencing (ChIP-seq) of histone H3K27 acetylation (H3K27ac), a mark of active enhancers (Hnisz et al., 2013), we first identified enhancer regions in *MYCN*-amplified Kelly and nonamplified SH-SY5Y cells (Figure S6A). In both cell types, a subset of H3K27ac-bound enhancers (Kelly 6.1%, SH-SY5Y, 5.4%) had a significantly higher signal than regular enhancers and were therefore classified as SEs (Lovén et al., 2013; Whyte et al., 2013) (Figure 6A). We also analyzed the occupancy of H3K4 monomethylation, a mark that colocalizes with H3K27ac in strong enhancers, noting a high concordance (94%–99%) with enhancers identified by H3K27ac analysis (Figures 6B and S6B).

The SEs in *MYCN*-amplified cells were either unique or disproportionately loaded with H3K27ac, with the largest SE being associated with the *MYCN* oncogene itself, having an H3K27Ac signal that was at least ten times greater than that of any other SE in either these or *MYCN*-nonamplified cells (Figures 6A and 6B).

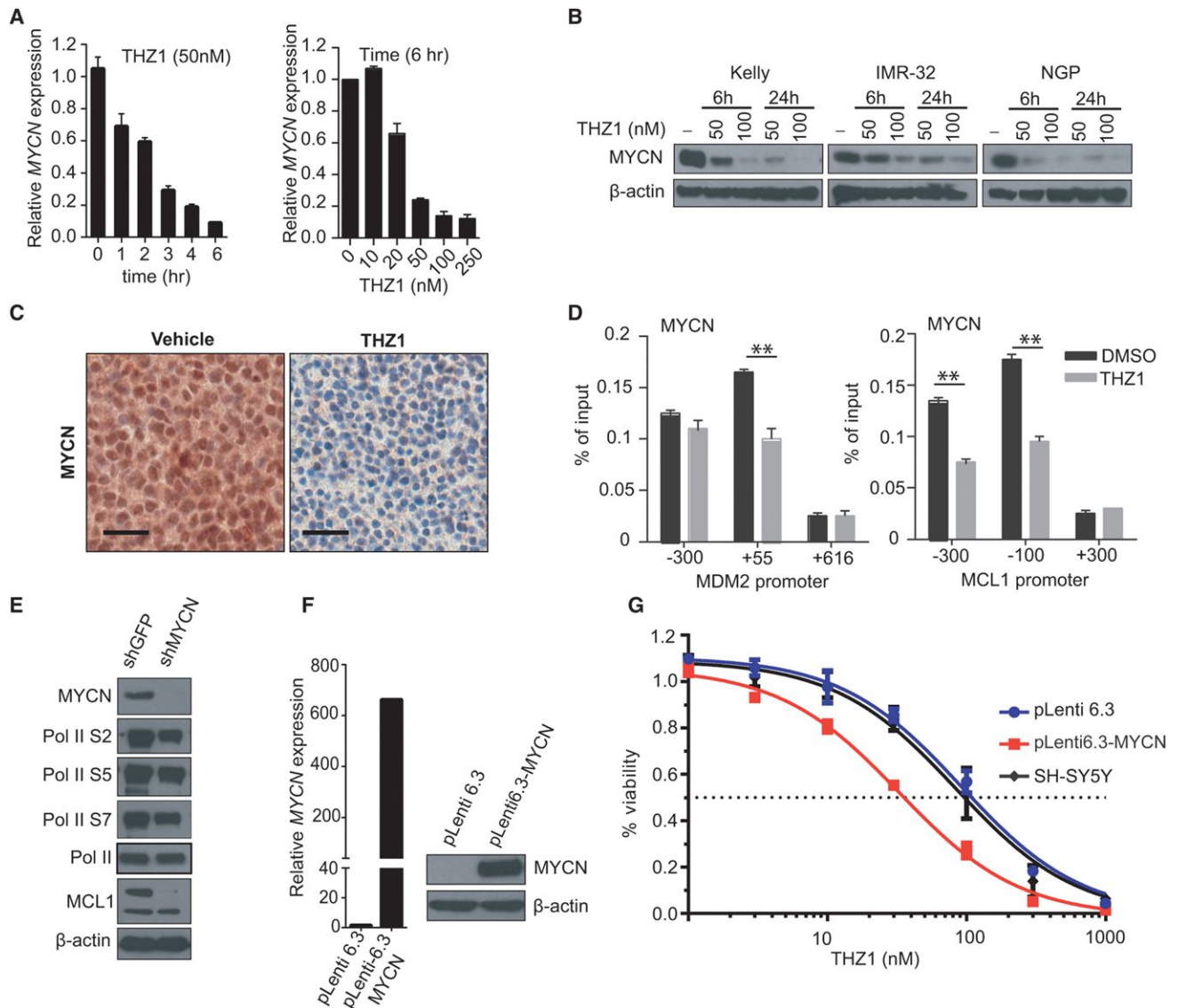


Figure 4. Cytotoxic Effects of THZ1 Are Mediated in Part through Inhibition of MYCN Expression

(A) Quantitative RT-PCR (qRT-PCR) analysis of *MYCN* RNA expression in *MYCN*-amplified Kelly cells treated with THZ1 50 nM for 0–6 hr (left) or 0–250 nM for 6 hr (right). Data normalized to β -actin are presented as mean \pm SD of three biological replicates.

(B) Immunoblot analysis of *MYCN* protein expression in *MYCN*-amplified NB cells treated with the indicated doses of THZ1 for 6 or 24 hr.

(C) IHC analysis of *MYCN* protein expression in *MYCN*-amplified human NB xenograft models treated with either DMSO or 10 mg/kg i.v. twice daily of THZ1 for 12 days. Scale bars represent 25 μ M.

(D) ChIP-qPCR analysis of *MYCN* binding at the promoters of candidate target genes in *MYCN*-amplified cells following treatment with THZ1, 100 nM for 3 hr. Mean \pm SD values for three replicate experiments are shown. ** $p < 0.01$ (Student's *t* test).

(E) Immunoblot analysis of the indicated proteins in *MYCN*-amplified NB cells expressing either an shRNA control (shGFP) or an shRNA directed against *MYCN* (shMYCN). Two different hairpins against *MYCN* [shMYCN(1) and shMYCN(3)] were used with similar results.

(F) qRT-PCR analysis of *MYCN* expression in SH-SY5Y *MYCN*-nonamplified cells engineered to express either a control vector (pLenti 6.3) or *MYCN* (pLenti 6.3-MYCN) (left). Immunoblot analysis of *MYCN* protein expression in these cells (right).

(G) Cell viability analysis of untransfected (SH-SY5Y), control vector-expressing (pLenti 6.3) or *MYCN*-expressing (pLenti 6.3-MYCN) *MYCN*-nonamplified NB cells treated for 72 hr with increasing doses of THZ1. Results are means \pm SD of three replicates.

See also Figure S4.

Because the *MYCN* SE in Kelly cells is contained within the *MYCN* amplicon, we questioned whether its SE status merely reflected the increased number of copies of the amplified gene. To investigate this possibility, we estimated the signal of a single

copy of *MYCN* and recalculated its rank within the list of enhancers. A single copy of the *MYCN* enhancer still fell within the group of SEs (ranked 250 out of 746), indicating that increased gene copy number does not account for the high SE

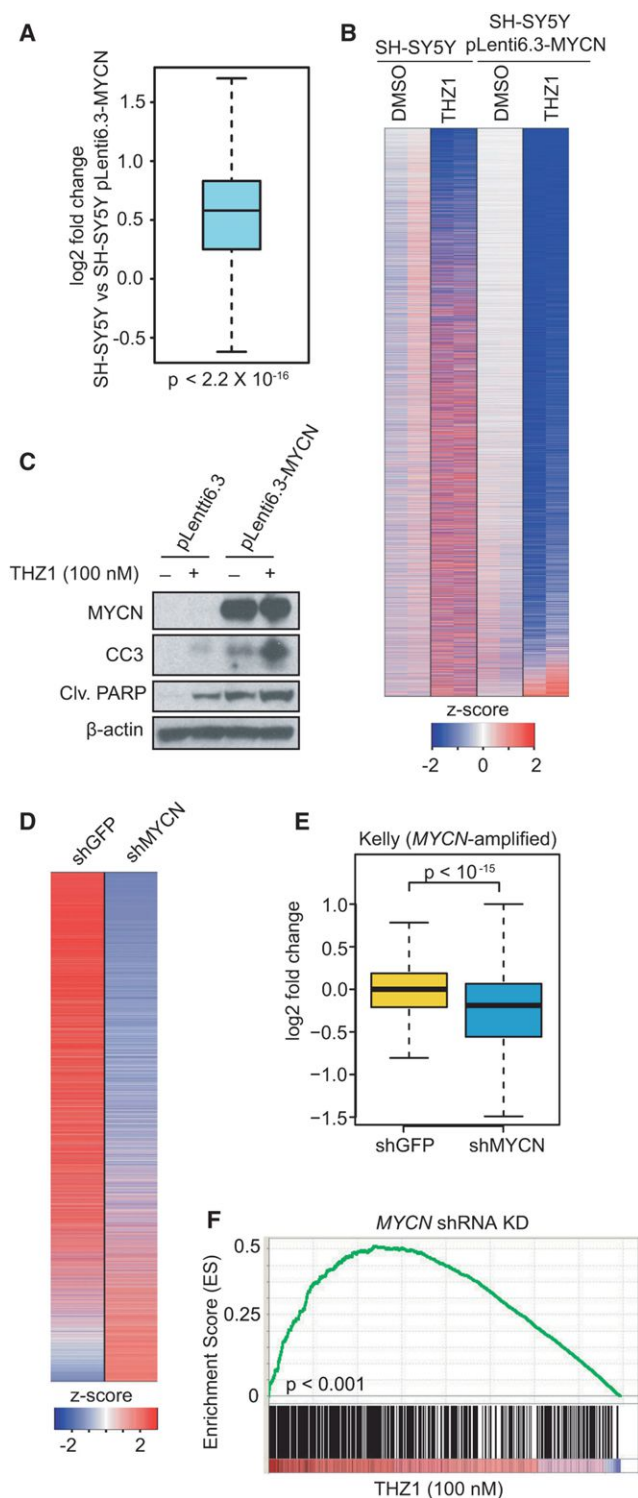


Figure 5. THZ1 Inhibits MYCN-Driven Transcriptional Amplification

(A) Box plot depicting the log₂ fold changes in actively transcribed genes in *MYCN*-nonamplified SH-SY5Y cells transduced with *MYCN* (pLenti 6.3-MYCN) compared with a control vector (pLenti 6.3). Box plot whiskers extend to 1.5 times the interquartile range ($n = 18,665$ expressed genes; $p < 2.2 \times 10^{-16}$, two-sided Mann-Whitney U test).

signal associated with *MYCN*. The majority of genes associated with the remaining top-ranked SEs were master transcription factors critical to sympathetic neuronal development and cell identity including *PHOX2B* (Pattyn et al., 1999; Stanke et al., 1999), *GATA2* (Tsarovina et al., 2004), *HAND2* (Howard et al., 2000), and *DBH* (Mercer et al., 1991) (Figures 6A and 6C). Another major oncogenic driver in NB, the receptor tyrosine kinase *ALK* (George et al., 2008), was associated with a unique top-ranked SE in *MYCN*-amplified cells (Figures 6A and 6C). In general, the SE domains associated with *MYCN* amplification correlated with pathways involved in the regulation of transcription, while those in nonamplified cells were primarily related to cell adhesion, invasion, metastasis and receptor tyrosine kinase pathways (Figures 6D and S6C). To ensure that the SEs in the cultured cells mirrored those of human NB, we analyzed H3K27ac occupancy in three primary tumors (Figures 6E and S6D). Consistent with our cell line data, the largest SE in *MYCN*-amplified samples NB#1 and NB#2 was associated with *MYCN* (Figures 6E and S6E). SEs associated with *PHOX2B* and *HAND2* were identified in all tumor samples (Figure 6E), in keeping with their lineage-specific role in sympathetic neuronal development.

To determine whether the occurrence of SEs correlates with increased Pol II occupancy, we used ChIP-seq analysis of Pol II and observed that the proximal promoter regions of SE-associated genes had higher levels of Pol II occupancy compared with regular enhancers in both *MYCN*-amplified and nonamplified cells, although the difference was much more striking in the former (Figure 7A). THZ1 treatment led to a striking reduction in Pol II binding at the promoter regions and gene bodies of SE-associated genes in *MYCN*-amplified but not nonamplified cells (Figure 7A). Pol II binding at regular enhancer-associated genes was not appreciably affected by THZ1 in either cell type (Figure 7A). The reduction in Pol II occupancy with THZ1 was observed at the transcription start and end sites of genes associated with the top-ranked SEs, including *MYCN*, suggesting that their expression was severely curtailed in these cells (Figure 7B). Indeed, RNA expression of SE-associated genes was significantly reduced after THZ1 treatment in *MYCN*-amplified cells (Figure 7C) but not of the same genes in *MYCN*-nonamplified cells (Figure 7C). Thus, not only do SEs in *MYCN*-amplified

(B) Heatmap of all gene expression values in *MYCN*-nonamplified SH-SY5Y cells versus SH-SY5Y cells overexpressing *MYCN* (pLenti 6.3-MYCN) following either THZ1 (100 nM for 6 hr) or DMSO treatment. Rows show Z scores calculated for each vector.

(C) Immunoblot analysis of MYCN, CC3, and cleaved PARP in SH-SY5Y *MYCN*-nonamplified cells engineered to express either a control vector (pLenti 6.3) or *MYCN* (pLenti 6.3-MYCN) following treatment with THZ1.

(D) Heatmap of differentially expressed genes in *MYCN*-amplified NB cells expressing a *MYCN* shRNA versus control shRNA. Rows show Z scores calculated for each vector.

(E) Box plots of log₂ fold changes in gene expression in *MYCN*-amplified NB cells expressing a *MYCN* shRNA versus a control shRNA. Box plot whiskers extend to 1.5 times the interquartile range ($n = 18,665$ expressed genes; $p < 10^{-15}$ for shMYCN versus shGFP, two-sided Mann-Whitney U test).

(F) GSEA plot depicting the correlation between the top 500 downregulated genes following THZ1 treatment and the rank-ordered genes that are differentially expressed after *MYCN* knockdown in *MYCN*-amplified NB cells. See also Figure S5.

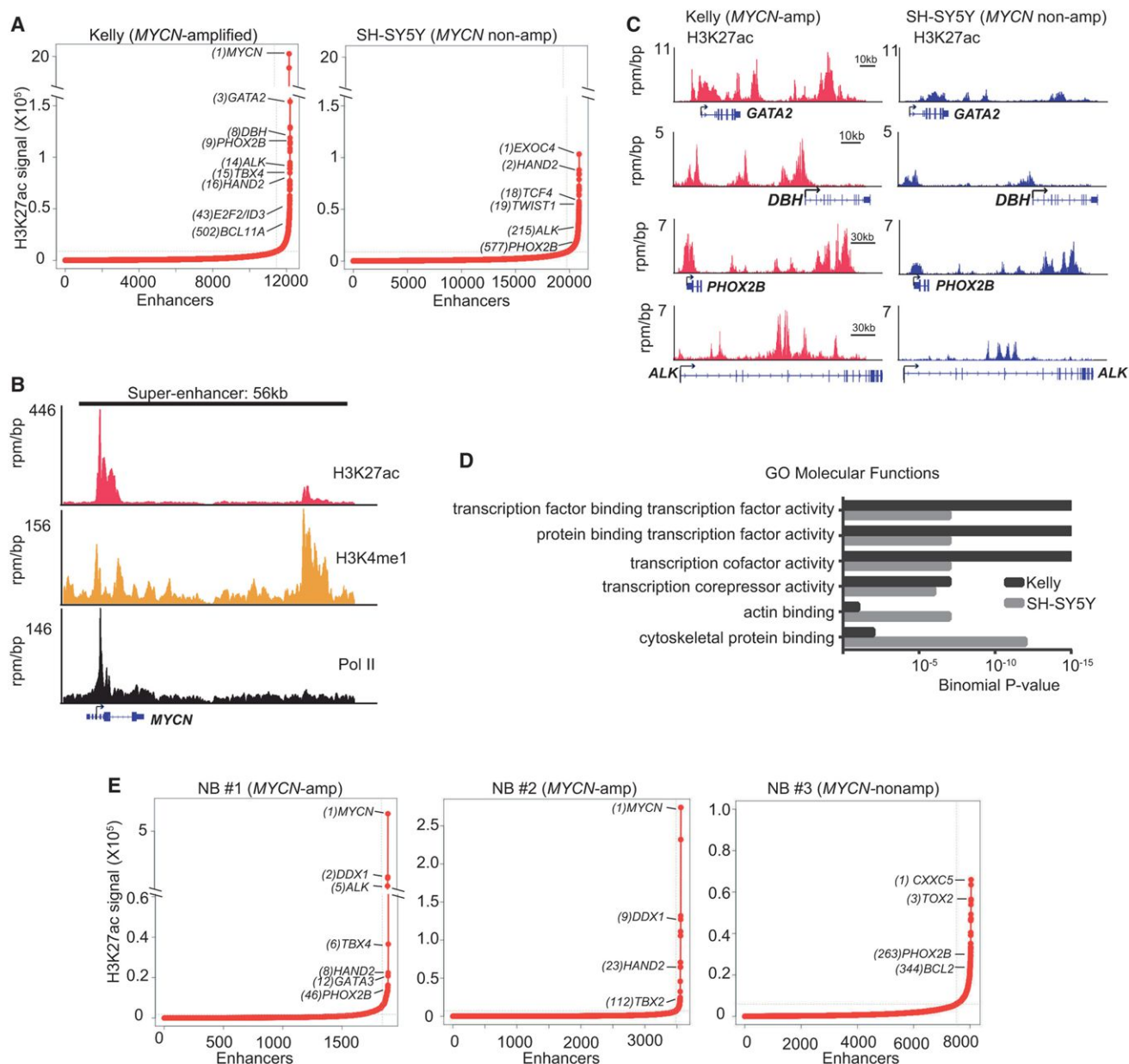


Figure 6. NB Cells Possess Unique Super-Enhancer Landscapes

(A) H3K27ac signal across enhancer regions for all enhancers in *MYCN*-amplified and nonamplified cells. SEs were defined as enhancers surpassing the threshold signal of 8,802 in both cell types. In Kelly and SH-SY5Y cells, 6.1% (746/12,200) and 5.4% (1,136/20,887) of the enhancers comprised 52% and 37%, respectively of all H3K27ac-bound enhancer signal and were classed as SEs.

(B) ChIP-seq profiles for H3K27ac, H3K4me1, and RNA Pol II binding at the *MYCN* SE gene locus in Kelly cells. The x axis shows genomic position and the y axis the signal of histone mark or Pol II binding in units of reads per million per base pair (rpm/bp). The gene model is depicted below and scale bars above the binding profiles.

(C) ChIP-seq profiles for H3K27ac binding at representative SE-associated gene loci in *MYCN*-amplified and nonamplified cells. The x and y axes are as described in (B).

(D) Gene Ontology (GO) molecular functions of SE-associated genes in *MYCN*-amplified (Kelly) and nonamplified (SH-SY5Y) cells identified using GREAT analysis (McLean et al., 2010).

(E) H3K27ac loading across enhancers in *MYCN*-amplified and nonamplified primary NB tumors. SEs were defined as having a threshold signal of 2,427 (NB#1), 5,933 (NB#2), and 5,952 (NB#3). In NB#1 and NB#2, 8.5% (164/1,920) and 3.0% (109/3,561), and in NB#3, 6.4% (522/8,040) of the enhancers were classified as SEs.

See also Figure S6.

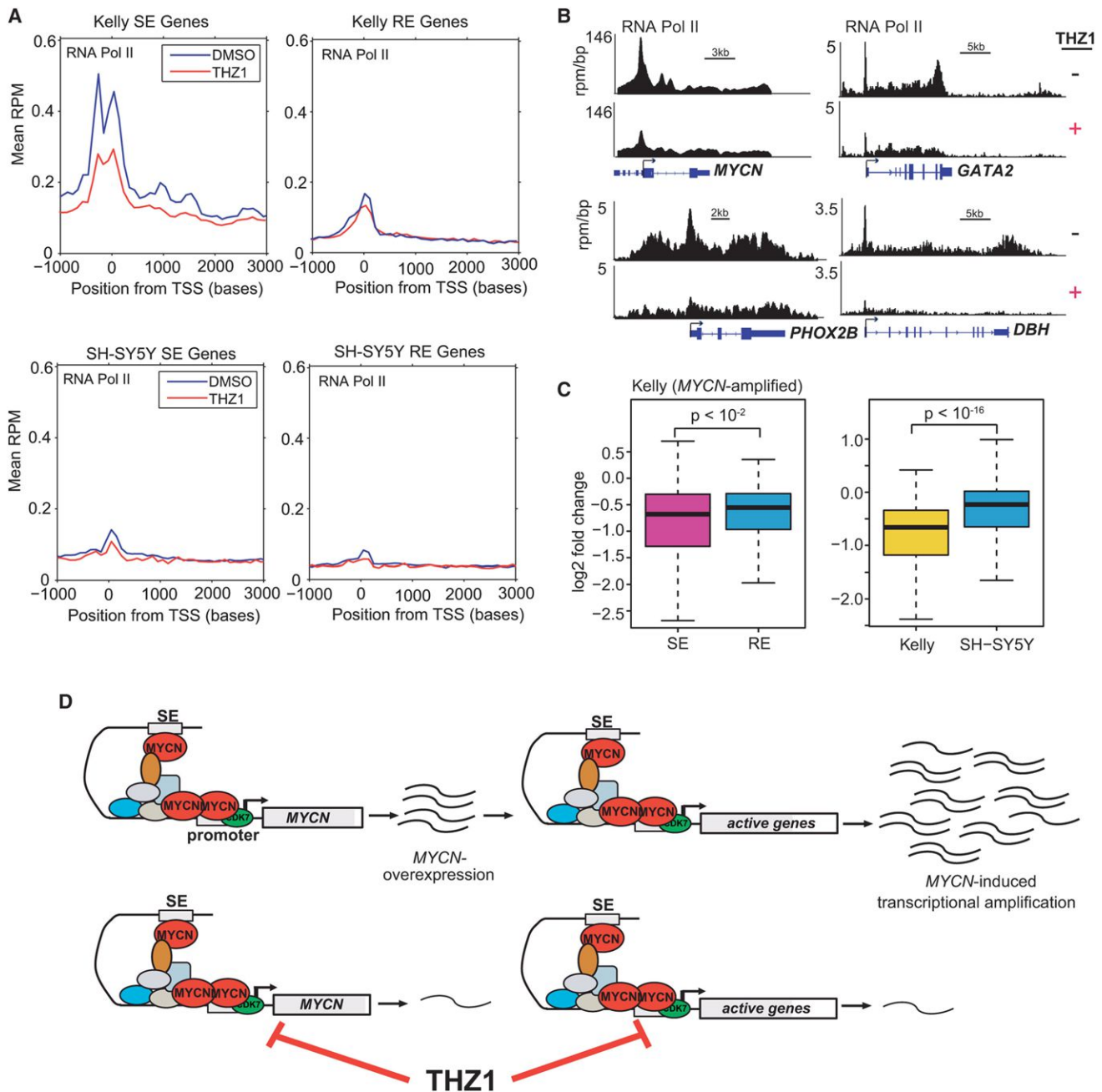


Figure 7. Sensitivity of MYCN-Amplified Cells to THZ1 Correlates with MYCN-Associated Super-Enhancers

(A) Metagenes representation of global Pol II ChIP-seq occupancy at SE- and RE-associated genes in cells with (Kelly) or without (SH-SY5Y) MYCN amplification treated with DMSO (blue) or THZ1 (red) (100 nM \times 3 hr). TSS, transcription start sites.

(B) Gene tracks of RNA Pol II binding density at representative SE-associated gene loci after DMSO or THZ1 treatment as in (A).

(C) Left: quartile box plots of log2 fold changes in the top 231 genes associated with SEs and regular enhancers (RE) in MYCN-amplified cells treated with THZ1 (100 nM \times 6 hr) versus DMSO. Box plot whiskers extend to 1.5 times the interquartile range ($n = 230$ SE; $n = 231$ RE; $p < 10^{-2}$, two-sided Mann-Whitney U test). Right: quartile box plots of log2 fold changes in gene transcripts associated with the top-ranked SEs unique to MYCN-amplified Kelly cells treated with THZ1 as in (A) versus DMSO, compared with the expression changes of the same genes in similarly treated SH-SY5Y nonamplified cells. Box plot whiskers extend to 1.5 times the interquartile range ($n = 673$; $p < 10^{-16}$, two-sided Mann-Whitney U test).

(D) Proposed mechanism for the action of THZ1 in MYCN-amplified NB. Oncogenic MYCN is regulated by super-enhancers leading to its high-level expression (left). Overexpressed MYCN invades the promoter and enhancer regions of all active genes, including itself, to induce global transcriptional upregulation (right). THZ1 targets the expression of both MYCN and MYCN-driven transcriptional amplification.

cells cause high-level expression of oncogenic drivers and genes that determine cell identity, but they also provide the basis for the selectivity of THZ1.

DISCUSSION

Here, we demonstrate that the genomically amplified *MYCN* oncogene, by promoting the development of SEs, causes upregulation of the active transcriptional program of NB cells, sensitizing them to inhibition of CDK7, a widely expressed regulator of transcription and cell-cycle transition. This effect suppresses essentially all active transcripts, especially those responsible for the uncontrolled proliferation that characterizes the cancer cell state, leading to potent and selective cytotoxicity while sparing normal cells. The high potency of our CDK7 inhibitor, THZ1, can be explained in part by its unprecedented mode of binding—the formation of a covalent bond with a unique cysteine residue outside the kinase domain—resulting in prolonged and irreversible CDK7 inactivation (Kwiatkowski et al., 2014). This feature contrasts with other transcriptional CDK inhibitors, including flavopiridol, roscovitine and CR-8 (Table S1), all of which rely on ATP-competitive modes of binding and whose effects on transcription would therefore be expected to be transient. The selectivity of THZ1 for *MYCN*-amplified cells can be attributed to the perturbation of SEs associated with deregulated *MYCN* and other oncogenic drivers.

Despite risk-based treatment strategies for *MYCN*-amplified NB, responses tend to be brief, with early relapses that are almost uniformly fatal (Matthay et al., 1999). Finding suitable ways to inhibit *MYCN*, or any *MYC* family member for that matter, has been notoriously difficult. A recent approach is the modulation of bromodomain and extraterminal (BET) domain coactivator proteins, such as BRD4, which associate with acetylated chromatin and promote transcriptional activation (Delmore et al., 2011; Lovén et al., 2013; Puissant et al., 2013). After successful testing of the BRD4 inhibitor JQ1 in *MYC*-dependent multiple myeloma models (Delmore et al., 2011), a positive correlation between *MYCN* amplification and sensitivity to JQ1 was demonstrated in the majority of *MYCN*-amplified NB models tested (Puissant et al., 2013). However, others reported growth inhibition of NB using the structurally distinct BRD4 inhibitor I-BET726, irrespective of *MYCN* amplification status or *MYC*/*MYCN* expression levels (Wyce et al., 2013). The intervention we describe has a number of features that would support the clinical development of covalent CDK7 inhibitors for use against *MYCN*-amplified NB and other *MYC*-driven cancers. These include the ability to achieve potent growth inhibition with relatively low doses of THZ1, the lack of off-target effects, and the absence of discernible toxicity in a mouse model of NB. THZ1 has been demonstrated to cross-react with CDK12/13 (Kwiatkowski et al., 2014); however, shRNA knockdown of these kinases in *MYCN*-amplified cells did not affect *MYCN* levels (Figure S1B). Furthermore, overexpression of *CDK7*^{C312S} restored *MYCN* levels to near amplified levels (Figure S4B). These results would suggest that loss of CDK7 activity dominates the selective effects of THZ1 in *MYCN*-amplified NB cells.

Multiple studies argue that *MYC* is a global transcriptional amplifier (Lin et al., 2012; Lovén et al., 2012; Nie et al., 2012;

Schuhmacher and Eick, 2013), although a recent report has suggested that a portion of this amplification is indirect (Sabò et al., 2014). We believe the transcriptional amplification effects of *MYC* are direct at the vast majority of active genes and that differences in interpretation of direct versus indirect effects are due to differences in interpretation of ChIP-seq data thresholds. We demonstrate here that deregulated *MYCN* functions as a transcriptional amplifier of the cell's existing gene expression program rather than a specific target gene set. This effect was not detectable in earlier data sets of *MYCN*-amplified tumor samples because of a limitation of microarray analysis, which in conventional approaches assumes similar levels of total RNA in all the samples (Lovén et al., 2012).

Although in principle, blockade of CDK7 function would be expected to inhibit the expression of any gene poised to undergo transcription, our direct demonstration that manipulation of *MYCN* levels can modulate the sensitivity to THZ1 supports the idea that *MYC*-induced transcriptional amplification is crucial to THZ1 activity. This transcriptional dependency was illustrated by the massive transcriptional shutdown in *MYCN*-amplified cells with relatively low doses of THZ1 (Figure 3). By contrast, although CDK transcripts were downregulated by THZ1 in *MYCN*-nonamplified cells, this effect did not produce the same profound impact seen in *MYCN*-amplified cells, providing further evidence that aberrant *MYCN* expression and the resultant global amplification of transcription are necessary for cell sensitivity to THZ1. This critical dependency on transcriptional amplification would account for the selective killing of *MYCN*-amplified cells by THZ1 without inducing toxic side effects in mice. The unique SE landscapes of *MYCN*-amplified cells also dictate their response to THZ1. Indeed, genes involved in the oncogenic state—such as *MYCN* and *ALK*, and those that specify sympathetic neuronal cell identity including *PHOX2B* and *HAND2*—were associated with SEs and were far more susceptible to THZ1 than either their counterparts in *MYCN*-nonamplified cells or genes driven by regular enhancers. This selectivity of THZ1 for *MYCN*-amplified cells could stem from its inhibition of the increased Pol II activity at SE-associated genes, leading to downregulation of *MYCN* expression as well as that of the entire transcriptional program of the cell.

Our data suggest that the cytotoxicity induced by THZ1 in *MYCN*-driven NB cells is mediated by inhibition of both *MYCN* expression and *MYCN*-stimulated global transcriptional amplification (Figure 7D). Oncogenic *MYCN* invades the core promoters and enhancer regions of actively transcribed genes, resulting in the formation of SEs at key genes underlying the cancer cell state, including *MYCN*, and ultimately in transcriptional amplification. THZ1, by irreversibly inhibiting CDK7, blocks transcription of *MYCN* (Figure 4) as well as *MYCN*-induced amplification of global gene transcription. The latter conclusion is based on the widespread inhibition of transcription in *MYCN*-amplified cells (Figure 3), the highly significant correlation between the transcripts that were downregulated by THZ1 and those resulting from genetic depletion of *MYCN*, and the fact that ectopic expression of *MYCN* in non-*MYCN*-overexpressing cells sensitizes these cells to widespread transcriptional inhibition by THZ1 (Figures 4 and 5). Together, these two processes effectively deprive tumor cells of the proliferative signals to which

they have become addicted, leading to apoptotic death. Nevertheless, this mechanism is not necessarily specific to deregulated *MYCN* or *MYC*. Rather, we suggest that inhibition of CDK7 or perhaps other transcriptional CDKs would be effective in any context in which tumor cells rely on high-level expression of one or more oncogenic transcription factors for their aberrant growth and survival. However, the net inhibitory effect is likely to be much more profound in cancers that depend on *MYC* (or *MYCN*) as the oncogenic driver.

In conclusion, we use NB and a newly developed transcriptional CDK inhibitor to demonstrate the potency and selectivity of a potential treatment strategy that targets global transcriptional amplification in *MYC*-driven tumors. Our results suggest that the mechanism by which oncogenic *MYC* gains control over myriad cellular processes to induce tumor formation also exposes a vulnerability that can be exploited therapeutically. By using THZ1 to disable essential components of the cancer cells' transcriptional machinery, we show that widespread suppression of transcription in *MYC*-dependent cancers is not only feasible, but also spares normal cells from toxicity.

EXPERIMENTAL PROCEDURES

Cell Lines

Human NB cell lines (from the Children's Oncology Group) and Raji and Daudi cells (provided by D. Sabatini, Massachusetts Institute of Technology) were grown in RPMI-1640, 10% fetal bovine serum (FBS), and 1% penicillin/streptomycin (pen/strep; Invitrogen). H262-BT111 primitive neuroectodermal tumor cells (provided by K. Ligon, Dana-Farber Cancer Institute [DFCI]) were grown in neuronal stem cell (NSC) media with heparin, NSC supplement (Stem Cell Technologies), 1% pen/strep, erythrocyte growth factor (EMD Millipore), and fibroblast growth factor (Life Technologies). NIH 3T3 cells (purchased from ATCC) and B6-MEFs (provided by A. Kung, Columbia University) were grown in DMEM with 10% FBS and 1% pen/strep.

Animal Studies

These experiments were performed with approval from the Institutional Animal Care and Use Committee of the DFCI. Full details are given in Extended Experimental Procedures.

Synthetic RNA Spike-In and Microarray Analysis

RNA samples were prepared as previously described (Lovén et al., 2012). Total RNA was spiked-in with RNA Spike-In Mix (Ambion), treated with DNA-free DNase I (Ambion), analyzed on Agilent 2100 Bioanalyzer (Agilent Technologies) for integrity, and hybridized to Affymetrix GeneChip PrimeView Human Gene Expression arrays (Affymetrix). Data analysis is described in Extended Experimental Procedures.

ChIP-Seq

ChIP was carried out as previously described (Lee et al., 2006) with minor changes described in the Extended Experimental Procedures. The following antibodies were used: RNAPII (Santa Cruz, sc-899), H3K4me1 (Abcam, ab8895), and H3K27ac (Abcam, ab4729).

Data Analysis

All ChIP-seq data were aligned using the short-read aligner Bowtie (version 0.12.7) (Langmead et al., 2009) to build version GRCh37/HG19 of the human genome. To visualize ChIP-seq tracks, reads were extended by 160 bases, converted into tdf files using igvtools (version 2.2.1) and visualized in IGV (Robinson et al., 2011). ChIP-seq peaks were detected using a peak-finding algorithm, MACS version 1.4.2 (Zhang et al., 2008) with the default p value threshold of enrichment of 1×10^{-5} used for all data sets. Active enhancers, ranked according to the magnitude of the H3K27ac signal, were defined as re-

gions of ChIP-seq enrichment for H3K27ac and H3K4me1 outside of promoters. To identify super-enhancers, we employed the ROSE algorithm (https://bitbucket.org/young_computation/rose) (Lovén et al., 2013; Whyte et al., 2013) to rank the enhancers that were identified using MACS in which peaks within 12.5 kb of one another were stitched together and these stitched enhancers were ranked by their difference in H3K27ac signal versus input signal. Further details are given in the Extended Experimental Procedures.

ACCESSION NUMBERS

The data discussed in this publication have been deposited in NCBI's Gene Expression Omnibus and are accessible through GEO series accession number GSE62726.

SUPPLEMENTAL INFORMATION

Supplemental Information includes Extended Experimental Procedures, six figures, and one table and can be found with this article online at <http://dx.doi.org/10.1016/j.cell.2014.10.024>.

ACKNOWLEDGMENTS

We thank K. Ligon, D. Sabatini, W. Weiss, A. Kung, and the Children's Oncology Group for cell lines and plasmids and P. Baran for the dCA compound. We thank J.R. Gilbert and Takaomi Sanda for insightful discussions. This study was supported by NIH R01CA148688 (R.E.G.), R01CA148688S1 (E.C.), R01CA179483-01 (N.S.G. and T.H.Z.), CA109901, HG002668 (R.A.Y., N.K., B.A.), and R21HG006778 (G.C.Y.); American Cancer Society (RSG-12-247-TBG to R.E.G.); Department of Defense (PR120741A to R.E.G.); and Friends for Life Neuroblastoma Foundation (N.S.G. and R.E.G.). THZ1 has been licensed to Syros Pharmaceuticals for clinical development. N.S.G. and R.A.Y. are scientific founders of Syros Pharmaceuticals.

Received: May 29, 2014

Revised: August 18, 2014

Accepted: September 24, 2014

Published: November 6, 2014

REFERENCES

- Arvanitis, C., and Felsher, D.W. (2006). Conditional transgenic models define how MYC initiates and maintains tumorigenesis. *Semin. Cancer Biol.* 16, 313–317.
- Boon, K., Caron, H.N., van Asperen, R., Valentijn, L., Hermus, M.C., van Sluis, P., Roobeek, I., Weis, I., Voûte, P.A., Schwab, M., and Versteeg, R. (2001). N-myc enhances the expression of a large set of genes functioning in ribosome biogenesis and protein synthesis. *EMBO J.* 20, 1383–1393.
- Brodeur, G.M., Seeger, R.C., Schwab, M., Varmus, H.E., and Bishop, J.M. (1984). Amplification of N-myc in untreated human neuroblastomas correlates with advanced disease stage. *Science* 224, 1121–1124.
- Chapuy, B., McKeown, M.R., Lin, C.Y., Monti, S., Roemer, M.G., Qi, J., Rahl, P.B., Sun, H.H., Yeda, K.T., Doench, J.G., et al. (2013). Discovery and characterization of super-enhancer-associated dependencies in diffuse large B cell lymphoma. *Cancer Cell* 24, 777–790.
- Cohn, S.L., Salwen, H., Quasney, M.W., Ikegaki, N., Cowan, J.M., Herst, C.V., Kennett, R.H., Rosen, S.T., DiGiuseppe, J.A., and Brodeur, G.M. (1990). Prolonged N-myc protein half-life in a neuroblastoma cell line lacking N-myc amplification. *Oncogene* 5, 1821–1827.
- Delmore, J.E., Issa, G.C., Lemieux, M.E., Rahl, P.B., Shi, J., Jacobs, H.M., Kastiris, E., Gilpatrick, T., Paranal, R.M., Qi, J., et al. (2011). BET bromodomain inhibition as a therapeutic strategy to target c-Myc. *Cell* 146, 904–917.
- Eilers, M., and Eisenman, R.N. (2008). Myc's broad reach. *Genes Dev.* 22, 2755–2766.
- Fisher, R.P., and Morgan, D.O. (1994). A novel cyclin associates with MO15/CDK7 to form the CDK-activating kinase. *Cell* 78, 713–724.

- Garriga, J., and Graña, X. (2004). Cellular control of gene expression by T-type cyclin/CDK9 complexes. *Gene* 337, 15–23.
- George, R.E., Sanda, T., Hanna, M., Fröhling, S., Luther, W., 2nd, Zhang, J., Ahn, Y., Zhou, W., London, W.B., McGrady, P., et al. (2008). Activating mutations in ALK provide a therapeutic target in neuroblastoma. *Nature* 455, 975–978.
- Glover-Cutter, K., Laroche, S., Erickson, B., Zhang, C., Shokat, K., Fisher, R.P., and Bentley, D.L. (2009). TFIIF-associated Cdk7 kinase functions in phosphorylation of C-terminal domain Ser7 residues, promoter-proximal pausing, and termination by RNA polymerase II. *Mol. Cell. Biol.* 29, 5455–5464.
- Hnisz, D., Abraham, B.J., Lee, T.I., Lau, A., Saint-André, V., Sigova, A.A., Hoke, H.A., and Young, R.A. (2013). Super-enhancers in the control of cell identity and disease. *Cell* 155, 934–947.
- Howard, M.J., Stanke, M., Schneider, C., Wu, X., and Rohrer, H. (2000). The transcription factor dHAND is a downstream effector of BMPs in sympathetic neuron specification. *Development* 127, 4073–4081.
- Kohl, N.E., Legouy, E., DePinho, R.A., Nisen, P.D., Smith, R.K., Gee, C.E., and Alt, F.W. (1986). Human N-myc is closely related in organization and nucleotide sequence to c-myc. *Nature* 319, 73–77.
- Kwiatkowski, N., Zhang, T., Rahl, P.B., Abraham, B.J., Reddy, J., Ficarro, S.B., Dastur, A., Amzallag, A., Ramaswamy, S., Tesar, B., et al. (2014). Targeting transcription regulation in cancer with a covalent CDK7 inhibitor. *Nature* 511, 616–620.
- Labisso, W.L., Wirth, M., Stojanovic, N., Stauber, R.H., Schnieke, A., Schmid, R.M., Krämer, O.H., Saur, D., and Schneider, G. (2012). MYC directs transcription of MCL1 and eIF4E genes to control sensitivity of gastric cancer cells toward HDAC inhibitors. *Cell Cycle* 11, 1593–1602.
- Lam, L.T., Pickeral, O.K., Peng, A.C., Rosenwald, A., Hurt, E.M., Giltman, J.M., Averett, L.M., Zhao, H., Davis, R.E., Sathyanarayanan, M., et al. (2001). Genomic-scale measurement of mRNA turnover and the mechanisms of action of the anti-cancer drug flavopiridol. *Genome Biol.* 2, RESEARCH0041.
- Langmead, B., Trapnell, C., Pop, M., and Salzberg, S.L. (2009). Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol.* 10, R25.
- Lapenna, S., and Giordano, A. (2009). Cell cycle kinases as therapeutic targets for cancer. *Nat. Rev. Drug Discov.* 8, 547–566.
- Laroche, S., Merrick, K.A., Terret, M.E., Wohlbold, L., Barboza, N.M., Zhang, C., Shokat, K.M., Jallepalli, P.V., and Fisher, R.P. (2007). Requirements for Cdk7 in the assembly of Cdk1/cyclin B and activation of Cdk2 revealed by chemical genetics in human cells. *Mol. Cell* 25, 839–850.
- Laroche, S., Amat, R., Glover-Cutter, K., Sansó, M., Zhang, C., Allen, J.J., Shokat, K.M., Bentley, D.L., and Fisher, R.P. (2012). Cyclin-dependent kinase control of the initiation-to-elongation switch of RNA polymerase II. *Nat. Struct. Mol. Biol.* 19, 1108–1115.
- Lee, T.I., Johnstone, S.E., and Young, R.A. (2006). Chromatin immunoprecipitation and microarray-based analysis of protein location. *Nat. Protoc.* 1, 729–748.
- Lin, C.Y., Lovén, J., Rahl, P.B., Paranal, R.M., Burge, C.B., Bradner, J.E., Lee, T.I., and Young, R.A. (2012). Transcriptional amplification in tumor cells with elevated c-Myc. *Cell* 151, 56–67.
- Lovén, J., Orlando, D.A., Sigova, A.A., Lin, C.Y., Rahl, P.B., Burge, C.B., Levins, D.L., Lee, T.I., and Young, R.A. (2012). Revisiting global gene expression analysis. *Cell* 151, 476–482.
- Lovén, J., Hoke, H.A., Lin, C.Y., Lau, A., Orlando, D.A., Vakoc, C.R., Bradner, J.E., Lee, T.I., and Young, R.A. (2013). Selective inhibition of tumor oncogenes by disruption of super-enhancers. *Cell* 153, 320–334.
- Malynn, B.A., de Alboran, I.M., O'Hagan, R.C., Bronson, R., Davidson, L., DePinho, R.A., and Alt, F.W. (2000). N-myc can functionally replace c-myc in murine development, cellular growth, and differentiation. *Genes Dev.* 14, 1390–1399.
- Matthay, K.K., Villablanca, J.G., Seeger, R.C., Stram, D.O., Harris, R.E., Ramsay, N.K., Swift, P., Shimada, H., Black, C.T., Brodeur, G.M., et al.; Children's Cancer Group (1999). Treatment of high-risk neuroblastoma with intensive chemotherapy, radiotherapy, autologous bone marrow transplantation, and 13-cis-retinoic acid. *N. Engl. J. Med.* 341, 1165–1173.
- McLean, C.Y., Bristor, D., Hiller, M., Clarke, S.L., Schaar, B.T., Lowe, C.B., Wenger, A.M., and Bejerano, G. (2010). GREAT improves functional interpretation of cis-regulatory regions. *Nat. Biotechnol.* 28, 495–501.
- Mercer, E.H., Hoyle, G.W., Kapur, R.P., Brinster, R.L., and Palmiter, R.D. (1991). The dopamine beta-hydroxylase gene promoter directs expression of *E. coli lacZ* to sympathetic and other neurons in adult transgenic mice. *Neuron* 7, 703–716.
- Nie, Z., Hu, G., Wei, G., Cui, K., Yamane, A., Resch, W., Wang, R., Green, D.R., Tessarollo, L., Casellas, R., et al. (2012). c-Myc is a universal amplifier of expressed genes in lymphocytes and embryonic stem cells. *Cell* 151, 68–79.
- Nishikura, K., Erikson, J., ar-Rushdi, A., Huebner, K., and Croce, C.M. (1985). The translocated c-myc oncogene of Raji Burkitt lymphoma cells is not expressed in human lymphoblastoid cells. *Proc. Natl. Acad. Sci. USA* 82, 2900–2904.
- Palancade, B., and Bensaude, O. (2003). Investigating RNA polymerase II carboxyl-terminal domain (CTD) phosphorylation. *FEBS* 270, 3859–3870.
- Pattyn, A., Morin, X., Cremer, H., Goridis, C., and Brunet, J.F. (1999). The homeobox gene *Phox2b* is essential for the development of autonomic neural crest derivatives. *Nature* 399, 366–370.
- Puissant, A., Frumm, S.M., Alexe, G., Bassil, C.F., Qi, J., Chanthery, Y.H., Nekritz, E.A., Zeid, R., Gustafson, W.C., Greninger, P., et al. (2013). Targeting MYCN in neuroblastoma by BET bromodomain inhibition. *Cancer Discov.* 3, 308–323.
- Robinson, J.T., Thorvaldsdóttir, H., Winckler, W., Guttman, M., Lander, E.S., Getz, G., and Mesirov, J.P. (2011). Integrative genomics viewer. *Nat. Biotechnol.* 29, 24–26.
- Rossignol, M., Kolb-Cheynel, I., and Egly, J.M. (1997). Substrate specificity of the cdk-activating kinase (CAK) is altered upon association with TFIIF. *EMBO J.* 16, 1628–1637.
- Sabò, A., Kress, T.R., Pelizzola, M., de Pretis, S., Gorski, M.M., Tesi, A., Morelli, M.J., Bora, P., Doni, M., Verrecchia, A., et al. (2014). Selective transcriptional regulation by Myc in cellular growth control and lymphomagenesis. *Nature* 511, 488–492.
- Schuhmacher, M., and Eick, D. (2013). Dose-dependent regulation of target gene expression and cell proliferation by c-Myc levels. *Transcription* 4, 192–197.
- Schwab, M., Alitalo, K., Klemm, K.H., Varmus, H.E., Bishop, J.M., Gilbert, F., Brodeur, G., Goldstein, M., and Trent, J. (1983). Amplified DNA with limited homology to myc cellular oncogene is shared by human neuroblastoma cell lines and a neuroblastoma tumour. *Nature* 305, 245–248.
- Seeger, R.C., Brodeur, G.M., Sather, H., Dalton, A., Siegel, S.E., Wong, K.Y., and Hammond, D. (1985). Association of multiple copies of the N-myc oncogene with rapid progression of neuroblastomas. *N. Engl. J. Med.* 313, 1111–1116.
- Serizawa, H., Mäkelä, T.P., Conaway, J.W., Conaway, R.C., Weinberg, R.A., and Young, R.A. (1995). Association of Cdk-activating kinase subunits with transcription factor TFIIF. *Nature* 374, 280–282.
- Slack, A., Chen, Z., Tonelli, R., Pule, M., Hunt, L., Pession, A., and Shohet, J.M. (2005). The p53 regulatory gene MDM2 is a direct transcriptional target of MYCN in neuroblastoma. *Proc. Natl. Acad. Sci. USA* 102, 731–736.
- Soucek, L., Whitfield, J., Martins, C.P., Finch, A.J., Murphy, D.J., Sodir, N.M., Karnezis, A.N., Swigart, L.B., Nasi, S., and Evan, G.I. (2008). Modelling Myc inhibition as a cancer therapy. *Nature* 455, 679–683.
- Stanke, M., Junghans, D., Geissen, M., Goridis, C., Ernsberger, U., and Rohrer, H. (1999). The Phox2 homeodomain proteins are sufficient to promote the development of sympathetic neurons. *Development* 126, 4087–4094.
- Toyoshima, M., Howie, H.L., Imakura, M., Walsh, R.M., Annis, J.E., Chang, A.N., Frazier, J., Chau, B.N., Loboda, A., Linsley, P.S., et al. (2012). Functional

genomics identifies therapeutic targets for MYC-driven cancer. *Proc. Natl. Acad. Sci. USA* **109**, 9545–9550.

Tsarovina, K., Pattyn, A., Stubbusch, J., Müller, F., van der Wees, J., Schneider, C., Brunet, J.F., and Rohrer, H. (2004). Essential role of Gata transcription factors in sympathetic neuron development. *Development* **131**, 4775–4786.

Veronese, M.L., Ohta, M., Finan, J., Nowell, P.C., and Croce, C.M. (1995). Detection of myc translocations in lymphoma cells by fluorescence in situ hybridization with yeast artificial chromosomes. *Blood* **85**, 2132–2138.

Wasylishen, A.R., and Penn, L.Z. (2010). Myc: the beauty and the beast. *Genes Cancer* **1**, 532–541.

Whyte, W.A., Orlando, D.A., Hnisz, D., Abraham, B.J., Lin, C.Y., Kagey, M.H., Rahl, P.B., Lee, T.I., and Young, R.A. (2013). Master transcription factors and mediator establish super-enhancers at key cell identity genes. *Cell* **153**, 307–319.

Wyce, A., Ganji, G., Smitheman, K.N., Chung, C.W., Korenchuk, S., Bai, Y., Barbash, O., Le, B., Craggs, P.D., McCabe, M.T., et al. (2013). BET inhibition silences expression of MYCN and BCL2 and induces cytotoxicity in neuroblastoma tumor models. *PLoS ONE* **8**, e72967.

Zhang, Y., Liu, T., Meyer, C.A., Eeckhoutte, J., Johnson, D.S., Bernstein, B.E., Nusbaum, C., Myers, R.M., Brown, M., Li, W., and Liu, X.S. (2008). Model-based analysis of ChIP-Seq (MACS). *Genome Biol.* **9**, R137.

Circadian Enhancers Coordinate Multiple Phases of Rhythmic Gene Transcription In Vivo

Bin Fang,^{1,2} Logan J. Everett,^{1,2} Jennifer Jager,^{1,2} Erika Briggs,¹ Sean M. Armour,¹ Dan Feng,¹ Ankur Roy,¹ Zachary Gerhart-Hines,¹ Zheng Sun,¹ and Mitchell A. Lazar^{1,*}

¹Division of Endocrinology, Diabetes, and Metabolism, Department of Medicine, Department of Genetics, and The Institute for Diabetes, Obesity, and Metabolism, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA 19104, USA

²Co-first author

*Correspondence: lazar@mail.med.upenn.edu

<http://dx.doi.org/10.1016/j.cell.2014.10.022>

SUMMARY

Mammalian transcriptomes display complex circadian rhythms with multiple phases of gene expression that cannot be accounted for by current models of the molecular clock. We have determined the underlying mechanisms by measuring nascent RNA transcription around the clock in mouse liver. Unbiased examination of enhancer RNAs (eRNAs) that cluster in specific circadian phases identified functional enhancers driven by distinct transcription factors (TFs). We further identify on a global scale the components of the TF cistromes that function to orchestrate circadian gene expression. Integrated genomic analyses also revealed mechanisms by which a single circadian factor controls opposing transcriptional phases. These findings shed light on the diversity and specificity of TF function in the generation of multiple phases of circadian gene transcription in a mammalian organ.

INTRODUCTION

A substantial proportion of mammalian genes are expressed with a circadian rhythm driven by a cell autonomous molecular clock (Hughes et al., 2009; Miller et al., 2007; Panda et al., 2002). The clock mechanism involves a network of transcriptional-translational feedback loops comprised of core transcriptional activators BMAL1/CLOCK and two sets of repressors, PER/CRY (Reppert and Weaver, 2001; Takahashi et al., 2008) and Rev-erbs α and β (Bugge et al., 2012; Cho et al., 2012; Ripperger and Schibler, 2001). Under normal conditions, each cellular clock is synchronized by systemic cues and generates multiple phases of rhythmic output (Asher and Schibler, 2011; Dibner et al., 2010; Peek et al., 2012).

Although each circadian transcription factor (TF) binds DNA with genome-wide oscillation peaking at a specific time (Feng et al., 2011; Koike et al., 2012; Rey et al., 2011), binding of an individual circadian TF, e.g., BMAL1, has been reported at genes oscillating with a range of phases, many of which do

not correlate with the circadian regulator's binding phase (Menet et al., 2012). Moreover, genome-wide studies have revealed a substantial portion of circadian TF binding tens to hundreds of kilobases away from known transcription start sites (TSS) (Feng et al., 2011; Koike et al., 2012; Rey et al., 2011) and a high degree of overlap between core clock TFs with competing effects on circadian rhythms, such as BMAL1 and Rev-erb α (Cho et al., 2012; Koike et al., 2012). Furthermore, several clock output TFs have been suggested to generate transcriptional rhythms with delayed phase relative to BMAL1/CLOCK, but these mechanisms have not been explored genome-wide (Asher and Schibler, 2011). Thus, a fundamental question remains as to how the interaction of multiple regulators at the genome, particularly at distal enhancer elements, produces distinct phases of circadian transcriptional activity.

Here, we applied Global Run-On sequencing (GRO-seq) (Core et al., 2008; Wang et al., 2011) to mouse livers collected at multiple times of day to measure the circadian activity of enhancer regions based on enhancer RNA (eRNA) transcription (Hah et al., 2013; Kim et al., 2010). We identified thousands of oscillating enhancers with varying peak activity times, and in particular, we found that specific phases of oscillation are associated with distinct regulatory motifs and TF binding patterns. Our data suggest that specific phases of enhancer activity in vivo are achieved by a dominant regulator at each site, determined in part by sequence content, in contrast to combinatorial regulation models based primarily on synthetic in vitro models (Ukai-Tadenuma et al., 2008). Furthermore, we show that eRNA oscillations are highly predictive of the rhythmicity and phase of transcription at nearby genes, demonstrating a large-scale and previously unexplored role for distal regulatory elements in the generation of transcriptional rhythms. By combining circadian enhancer maps, transcription factor cistromes, and genetic ablation of *Rev-erb α* and *Clock*, we demonstrate that circadian eRNAs can be used to both identify the TFs coordinating specific phases of gene transcription and, importantly, uniquely distinguish the functional binding sites within a circadian TF cistrome. Thus, an integrative approach using multiple genomic techniques provides the most detailed and robust model to explain the generation and coordination of multiple phases of rhythm within a single tissue.

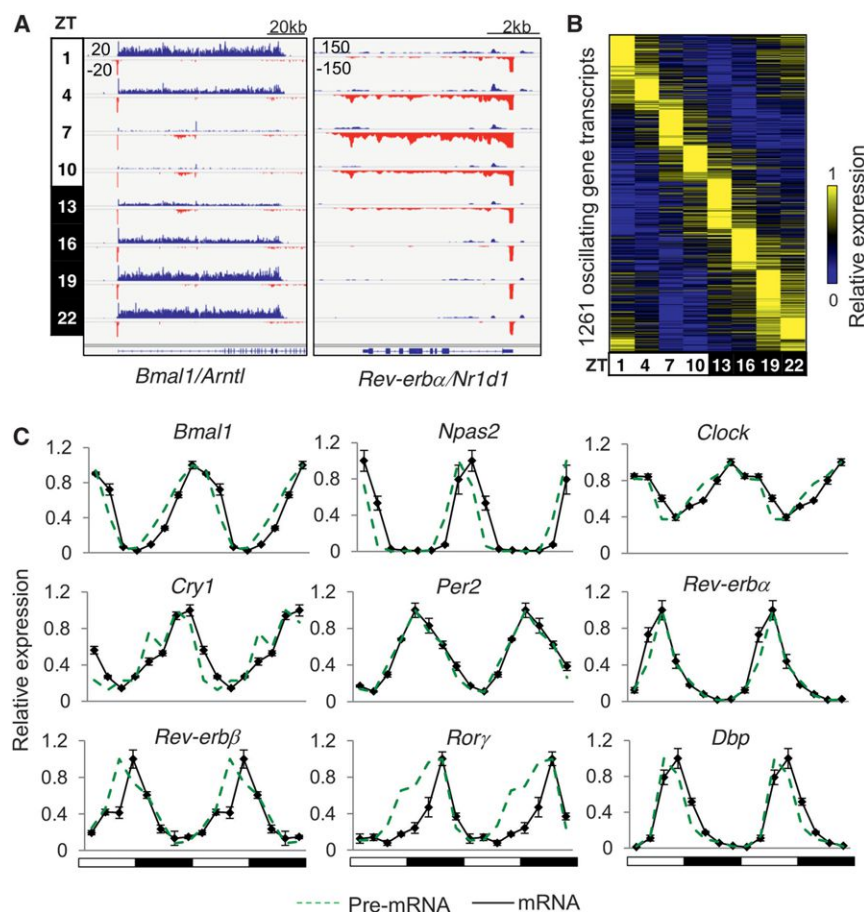


Figure 1. Circadian Transcription in Mouse Liver

(A) Genome browser view of nascent transcripts at *Bmal1/Arntl* and *Rev-erbα/Nr1d1* loci at eight time points. GRO-seq signals on the + and - strand are illustrated in blue and red, respectively. Y axis scale refers to the normalized tag count per million reads.

(B) Heat map of the relative transcription of 1,261 oscillating genes sorted by oscillation phase.

(C) Relative expression of pre-mRNA (green) and mRNA (black) determined by GRO-seq and RT-qPCR, respectively, throughout the day. Data are double plotted for better visualization. RT-qPCR data are expressed as the mean \pm SEM ($n = 3-4$ per time point) and normalized to the maximal expression of the day.

See also Figure S1 and Table S1.

regions in the vicinity of *Ppara* and *Cry2*, respectively (Figure 2A). To globally identify eRNA loci, we developed a pipeline to search for genomic locations producing bi- and unidirectional short RNA transcripts (Extended Experimental Procedures), which identified 19,086 high confidence de novo eRNA loci (>300 bp from TSS) (Table S2A). The average GRO-seq signal of de novo eRNAs showed a bimodal profile in both inter- and intragenic regions (Figure 2B). Analysis of public chromatin immunoprecipitation sequencing (ChIP-seq) data (Table

RESULTS

Circadian Transcription in Mouse Liver

GRO-seq was performed on mouse liver nuclei collected every three hours throughout a 24 hr light-dark cycle. Transcription of known circadian genes showed robust oscillation patterns, exemplified by *Bmal1* (*Arntl*) and *Rev-erbα* (*Nr1d1*) (Figure 1A). A total of 11,288 active gene transcripts were identified, of which 1,261 (11%) were transcribed with oscillating patterns (JTK_CYCLE [Hughes et al., 2010], $p < 0.01$, $21 \leq \text{period} (\tau) \leq 24$ hr, peak to trough ratio > 1.5) (Figure 1B; Table S1A available online). Rhythmic mRNA expression of known circadian genes determined by RT-quantitative PCR (RT-qPCR) was associated with their nascent transcription (Figure 1C), and biological replicates of GRO-seq samples at Zeitgeber Time (ZT) 10 and ZT22 showed a high degree of correlation (Pearson correlation coefficient, $r = 0.95$) (Figure S1A). In addition, genes oscillating in similar phases showed closely related biological functions (Figure S1B; Table S1B). Together, these results demonstrate the robustness of our data.

De Novo Identification of Circadian Liver Enhancer RNAs

Analysis of the liver GRO-seq data revealed eRNA transcription in both inter- and intragenic regions, exemplified by highlighted

S2B) from mouse liver suggested that de novo eRNA loci were enriched for other epigenomic features including H3K27ac, H3K4me1, DNase I hypersensitivity, and RNA polymerase II (Pol2) recruitment, consistent with the function of these sites as enhancers (Figure 2C). eRNA signals correlated with Pol2 occupancy and histone acetylation but not histone methylation (Figure S2A), consistent with earlier reports (Hah et al., 2013; Li et al., 2013; Wang et al., 2011) and in agreement with the notion that H3K4me1 and H3K27ac mark enhancer identity and activity, respectively (Creyghton et al., 2010).

To examine dynamics of eRNA transcription across the 24 hr cycle, eRNA transcripts were quantified using GRO-seq tag counts within ± 500 bp from the centers of eRNA loci. Remarkably, 5,724 (30%) of eRNAs were found to be transcribed in a circadian manner (JTK_CYCLE, $p < 0.05$, $21 \leq \text{period} (\tau) \leq 24$ hr, peak to trough ratio > 1.5) (Table S2C), and their relative expression peaked at different times of the day (Figure 2D). Based on their peak expression time (hereafter referred to as “phase”), circadian eRNAs were divided into eight groups (phase ZT0–ZT24, at 3 hr intervals), represented by eight colors in Figure 2D. Interestingly, circadian eRNAs were not evenly distributed across the eight phase groups. A total of 71% of circadian eRNAs oscillated with a phase between ZT18 and ZT3, whereas 29% of circadian eRNAs oscillated in other phases (Figure 2E; Table S2C). Examples of circadian eRNAs with phase

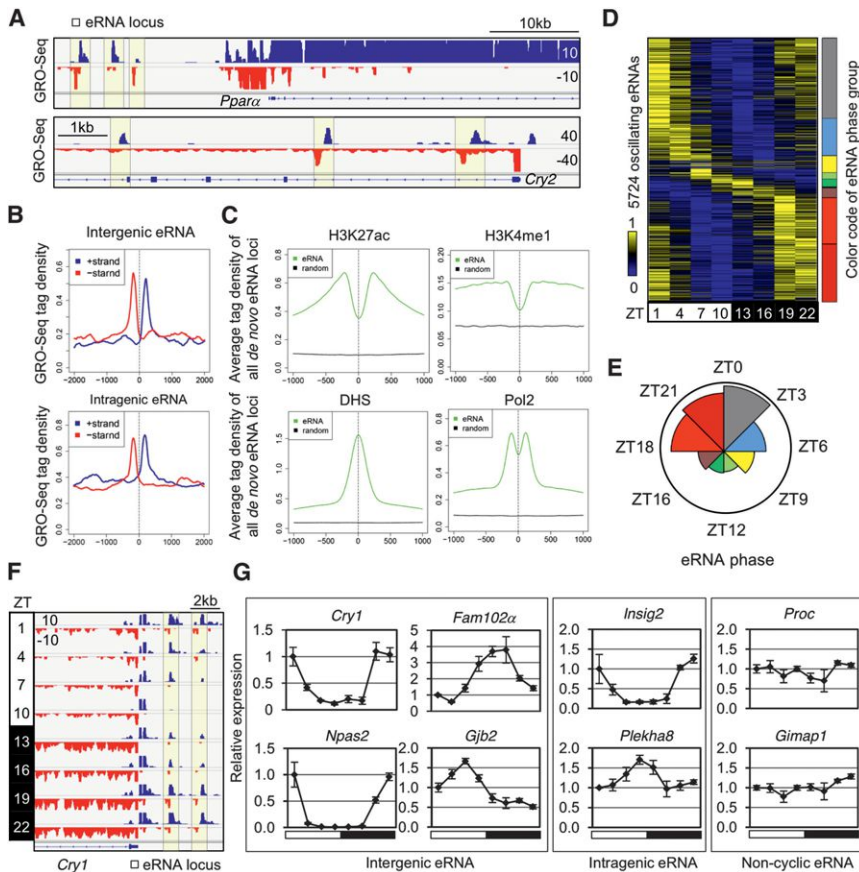


Figure 2. De Novo Identification of Circadian Liver Enhancer RNAs

(A) Genome browser view of intergenic (upper panel) and intragenic (lower panel) eRNAs (yellow boxes).

(B) GRO-seq tag densities in 4 kb windows surrounding de novo intergenic (upper panel) and intragenic (lower panel) eRNA loci are shown for the plus (blue) and minus (red) strand. Y axis shows average reads per 10 million reads (RPTM) per 10 bp bin.

(C) Average ChIP-seq tag densities of epigenetic marks in 2 kb window surrounding all de novo eRNA loci (prior to the selection of high confidence eRNAs) and matched control regions.

(D) Heat map of the relative transcription of oscillating eRNAs throughout the day. Color coding of eRNA population in eight phase groups (from ZT0 to ZT24, at 3 hr intervals) is shown on the right.

(E) Rose diagram showing the prevalence of eRNA loci in each phase group. For each wedge, the color corresponds to that in (D) and the area is proportional to the number of eRNAs in that group.

(F) Genome browser view of oscillating eRNAs at *Cry1* locus.

(G) RT-qPCR validation of circadian transcription for intergenic, intragenic, and noncyclic eRNAs at indicated gene loci. Data are expressed as mean \pm SEM ($n = 3-4$ per time point) and normalized to the first time point.

See also Figure S2 and Table S2.

ZT22 at the *Cry1* locus are shown in Figure 2F. eRNA transcripts oscillating in different phases were confirmed by RT-qPCR (Figure 2G) at selected intergenic and intragenic eRNA loci (Figure S2B). The unbalanced phase distribution of eRNAs agrees with the previous finding that histone acetylation, a reflection of enhancer activity, was globally high around ZT22 and low around ZT10 in the mouse liver (Feng et al., 2011). Moreover, the average H3K27ac level at eight groups of eRNA loci showed the same oscillatory pattern as the circadian eRNAs within each group (Figure S2C). Therefore, circadian eRNAs oscillate in diverse phases, suggesting that circadian enhancer activities are orchestrated by distinct mechanisms in liver.

Phase-Specific Transcription Factors at Circadian Enhancers

We have shown that gene body and eRNA transcription occur in multiple phases. As previous studies suggested correlated transcription of eRNA and nearby target genes (Core et al., 2008; Hah et al., 2013; Kim et al., 2010), we examined whether eRNA oscillations are related to circadian gene transcription. The expression of genes mapped closest to oscillating eRNAs (within 200 kb from TSS) showed rhythmic patterns in phase with eRNA expression (Figure 3A). Among all genes mapped to circadian eRNAs, 423 (34%) circadian gene transcripts were mapped to 1,124 (20%) circadian enhancers and oscillation phases between each enhancer-gene pair were highly correlated ($r = 0.9$)

(Figure S3A). This is likely an underestimate based on the stringent eRNA-gene mapping criteria and, indeed, if the analysis is not limited to the nearest gene, up to 76% of circadian genes in different phases have in-phase eRNAs (phase difference <3 hr between gene and eRNA) located within 200 kb of their TSSs. By contrast, for random genes this number is $\sim 10\%$ on average (hypergeometric test, $p < 0.001$) (Figure S3B). Together, these results suggest that circadian eRNAs predict rhythmic transcription of nearby genes and are likely to be functionally associated with circadian genes of the same phase.

Although gene body and eRNA transcription occur in multiple phases, the core clock oscillator in liver has only one peak and one trough in a 24 hr period (Koike et al., 2012). We considered the possibility that specific circadian TFs were responsible for the different phases of gene expression by driving the transcription of diversely phased eRNAs. To this end, we performed motif analysis on the eight groups of circadian enhancers using 500 bp windows centered on each eRNA locus (Figure S3C). First, candidate phase-specific TFs with the most enriched motifs in each enhancer group were selected by de novo motif mining (Table S3). Then, annotated motifs of candidate TFs were used to quantify the motif enrichment in each enhancer group, revealing four major types of motifs specifically enriched in six enhancer groups (Figure 3B). Specifically, an E-box motif was the most enriched at circadian eRNA loci in phase ZT6–ZT9, coincident with the peak of BMAL1 binding to the genome

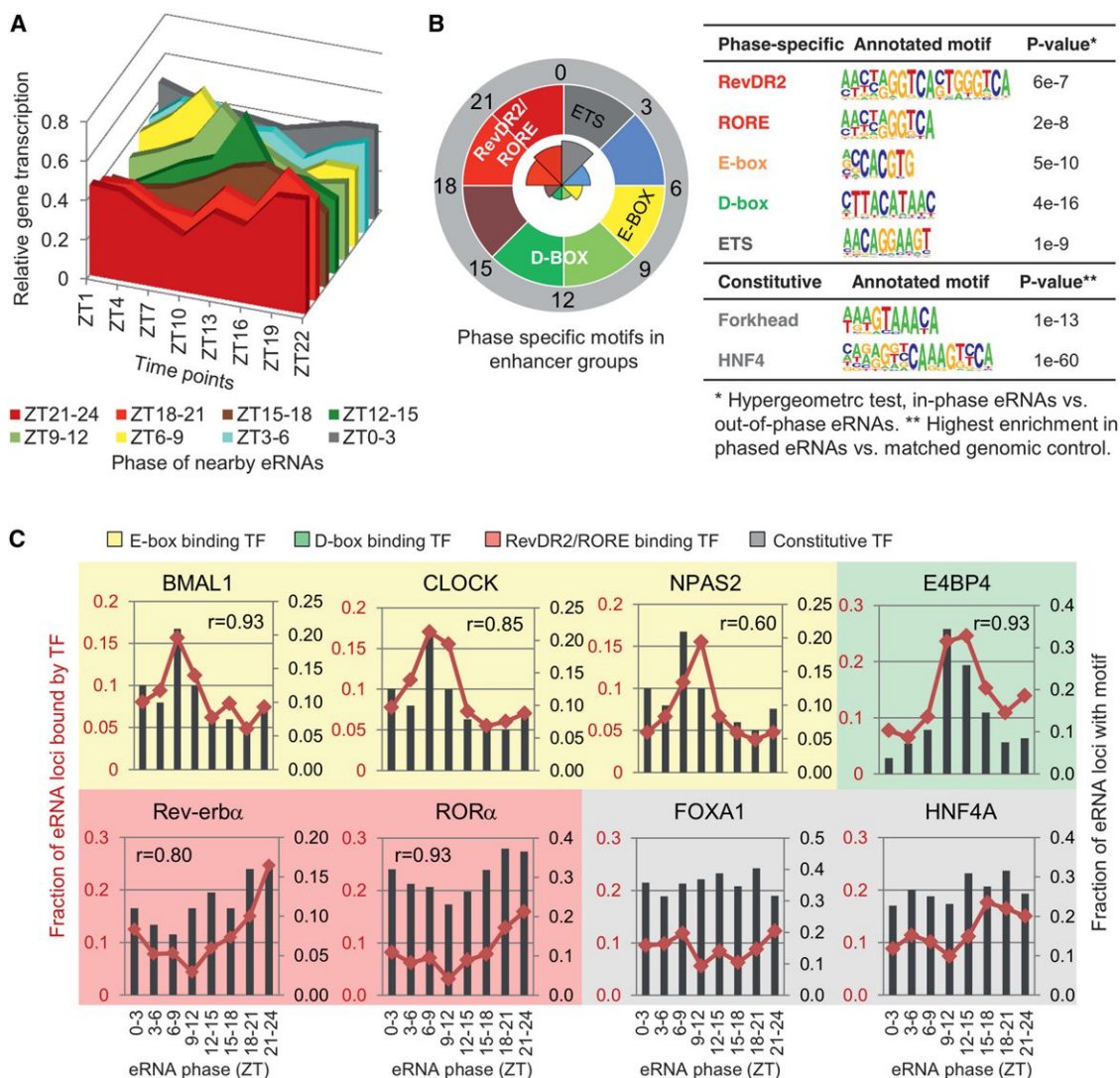


Figure 3. Phase-Specific Transcription Factors at Circadian Enhancers

(A) Relative transcription of genes closest to oscillating eRNAs (within 200 kb of TSS).

(B) Motifs specifically enriched in each eRNA group are labeled in the clock diagram on the left. Position weight matrix (PWM) of each motif and its best enrichment p value in assigned groups are shown in the table on the right.

(C) Correlation of motif occurrence and TF binding in eight eRNA phase groups. In each plot, the red dots represent the fraction of eRNA loci bound by the indicated TF (top 3,000 ChIP-seq peaks), and black bars represent the fraction of eRNA loci containing the corresponding motif. Correlation coefficient r is shown for phase-specific motifs. TFs recognizing different types of motifs are grouped in colored boxes corresponding to those used for eRNA phases.

See also Figure S3 and Table S3.

(Koike et al., 2012; Rey et al., 2011; Ripperger and Schibler, 2006). However, although BMAL1/CLOCK has been previously linked to circadian gene regulation in liver, the ZT6–ZT9 eRNAs comprised only ~6% of circadian enhancers, consistent with an earlier study in which only ~5% of total circadian genes were transcribed in phase with nearby BMAL1 binding (Menet et al., 2012).

We also discovered that a D-box motif, recognized by PAR-bZIP proteins including DBP, TEF, HLF, and E4BP4 (Cowell et al., 1992; Li and Hunger, 2001; Mitsui et al., 2001), was the most enriched motif at phase ZT9–ZT15 eRNA loci (Figure 3B),

coinciding with the phase of known target genes for these TFs (Gachon et al., 2006). Moreover, the RevDR2 and RORE motifs, bound by Rev-erb α/β (Harding and Lazar, 1995) and ROR α/γ (Giguère et al., 1994), were the top motifs at eRNA loci with the most common phase, ZT18–ZT24 (Figure 3B), coinciding with the trough of repression by Rev-erb α (Bugge et al., 2012; Feng et al., 2011). By contrast, motifs characteristic of ETS binding sites were highly enriched in the phase ZT0–ZT3 enhancers, implying a potential role of ETS proteins in the circadian regulation of transcripts with this phase (Figure 3B). In addition to these phase-specific motifs, constitutively enriched motifs in all

enhancer groups were identified, most prominently the Forkhead and HNF4 motifs (Figure 3B).

We tested whether the motif enrichment in a given eRNA group was predictive of TF binding by overlapping each group of circadian eRNAs with TF cistromes determined by ChIP-seq. Specifically, we analyzed previously published cistrome data for core clock TFs (Feng et al., 2011; Koike et al., 2012) and performed additional ChIP-seq experiments for E4BP4 and ROR α . To minimize the effects of variable ChIP-seq quality in different studies, only the 3,000 strongest ChIP-seq peaks for each TF were used in the analysis. Notably, the genomic binding sites of E-box-binding factors BMAL1, CLOCK, and NPAS2 were enriched at eRNAs with phase ZT6–ZT9 (Figure 3C), where de novo analysis implicated the E-box motif. Similarly, genomic binding of Rev-erb α and ROR α was enriched at eRNAs whose transcription peaked at ZT21–ZT24 (Figure 3C), where the RevDR2 and RORE motifs were most prominent. Also consistent with the bioinformatic predictions, the D-box binding factor E4BP4 bound most commonly at eRNAs with phase ZT9–ZT15 (Figure 3C). By contrast, binding of FOXA1 and HNF4A, whose motifs were equally enriched in all eRNA groups, did not display a preference for eRNA loci of a specific phase (Figure 3C). Thus, the regulatory activities of six TFs coincide with the rhythmic eRNA expression in the enhancer group at which they were enriched. These data strongly suggest that TFs bound specifically at each enhancer group are potential drivers of their circadian transcription and enhancer activities.

Phase Correlation between eRNA and Gene Body Transcription Marks Functional Enhancers of Circadian Genes

We next considered whether the specific TFs found to bind at circadian enhancers were driving transcription of nearby in-phase genes, focusing on the most common circadian enhancers (phase ZT18–ZT24). Within 200 kb of 325 circadian genes in phase ZT18–ZT24, 539 neighboring eRNA loci showed circadian eRNA transcription in phase ZT18–ZT24 (“correlated enhancers”), while 857 eRNA loci did not produce correlated eRNA transcription (“noncorrelated enhancers,” eRNA expression ZT22/ZT10 < 1.5) (Figure 4A).

Correlated enhancers showed higher enrichment of the RevDR2 and RORE motifs in comparison to noncorrelated enhancers (Figure 4B). Notably, relative enrichment of the RevDR2 motif, which is a preferential binding site for Rev-erb α (Harding and Lazar, 1995; Zhao et al., 1998) was 2-fold higher than that of the RORE motif shared by Rev-erb α and ROR α (Giguère et al., 1994), suggesting that Rev-erb α may play a more important role in regulating the correlated enhancers. ChIP-seq tag densities of Rev-erb α and its corepressor HDAC3 were dramatically stronger at correlated enhancers than at noncorrelated enhancers (Figure 4C), supporting the idea that the correlated enhancers in phase ZT18–ZT24 were controlled by Rev-erb α . To test this hypothesis, GRO-seq was performed on livers from mice genetically lacking Rev-erb α (Rev-erb α ^{−/−}) at ZT10, when Rev-erb α levels normally peak and maximally repress histone acetylation and gene transcription (Feng et al., 2011). Indeed, eRNA signals at the correlated enhancers were markedly derepressed in Rev-erb α ^{−/−} mice, while no such change was seen

at the noncorrelated enhancers (Figure 4D). Similar results were obtained at both inter- and intragenic enhancers (Figure S4). Importantly, gene body transcription that normally peaked at ZT18–ZT24 was also extensively derepressed in Rev-erb α ^{−/−} mice at ZT10 (Figure 4E), indicating these genes are direct targets of Rev-erb α . Together, these results demonstrate that eRNAs in phase ZT18–ZT24 mark functional Rev-erb α binding sites that regulate neighboring target genes with correlated phase. Conversely, noncorrelated enhancers are not bound by Rev-erb α and do not control Rev-erb α target genes.

Circadian eRNAs Reveal the Functional Rev-erb α Cistrome at Oscillating Genes

The findings to this point demonstrate that Rev-erb α regulates circadian genes in phase ZT18–ZT24 via enhancers oscillating in phase with gene body transcription. However, these enhancers account for only a small fraction of the complete Rev-erb α cistrome (Feng et al., 2011). We therefore considered whether circadian eRNAs in phase ZT18–ZT24 uniquely mark the functional subset of Rev-erb α binding sites controlling circadian genes in liver. To test this, Rev-erb α sites near circadian genes were divided into three groups, of which 887 (33%) overlapped de novo eRNA loci, 347 (13%) were found at TSSs of circadian genes (within 300 bp), and the remaining 1,455 (54%) were not associated with detectable eRNA transcription (Figure 5A). Of the eRNAs transcribed at Rev-erb α binding sites, 30% peaked at ZT18–ZT24, while 19% peaked in other phases, and 51% were constitutively expressed eRNA and did not oscillate (Figure 5A).

Rev-erb α and its corepressor HDAC3 bound more strongly at sites producing ZT18–ZT24 eRNAs than at other types of binding sites (Figure 5B), resulting in a marked decrease in histone H3K9 acetylation from ZT22 to ZT10 (Figure S5). To directly assess the functionality of Rev-erb α binding on individual gene expression, we constructed a list of high confidence target genes whose nascent and mature transcripts were derepressed in Rev-erb α ^{−/−} livers at ZT10 compared to wild-type (WT) (Tables S4A–S4C). The enrichment of derepressed circadian genes in Rev-erb α ^{−/−} mice was >3-fold higher near Rev-erb α sites producing ZT18–ZT24 eRNAs, compared to other Rev-erb α sites (Figure 5C), suggesting that ZT18–ZT24 eRNAs mark functional Rev-erb α sites. Moreover, circadian genes with phase around ZT21–ZT24 were highly enriched for derepression in Rev-erb α ^{−/−} mice (Figure 5D), consistent with the enrichment of circadian eRNAs in this phase. Together, these data strongly suggest that only a subset of the Rev-erb α cistrome associated with antiphase eRNAs is functional in controlling circadian gene transcription.

eRNA Analysis Identifies E4BP4 as a Key Mediator of Gene Activation by Rev-erb α

While eRNAs clearly delineate the functional Rev-erb α cistrome responsible for direct transcriptional repression, there remains a substantial set of genes paradoxically downregulated at ZT10 in Rev-erb α ^{−/−} mouse livers, which cannot be explained through direct regulation by Rev-erb α . To identify factors mediating this opposing effect on gene transcription, we constructed a list of high confidence target genes whose nascent and mature transcript levels were decreased in Rev-erb α ^{−/−} livers at ZT10

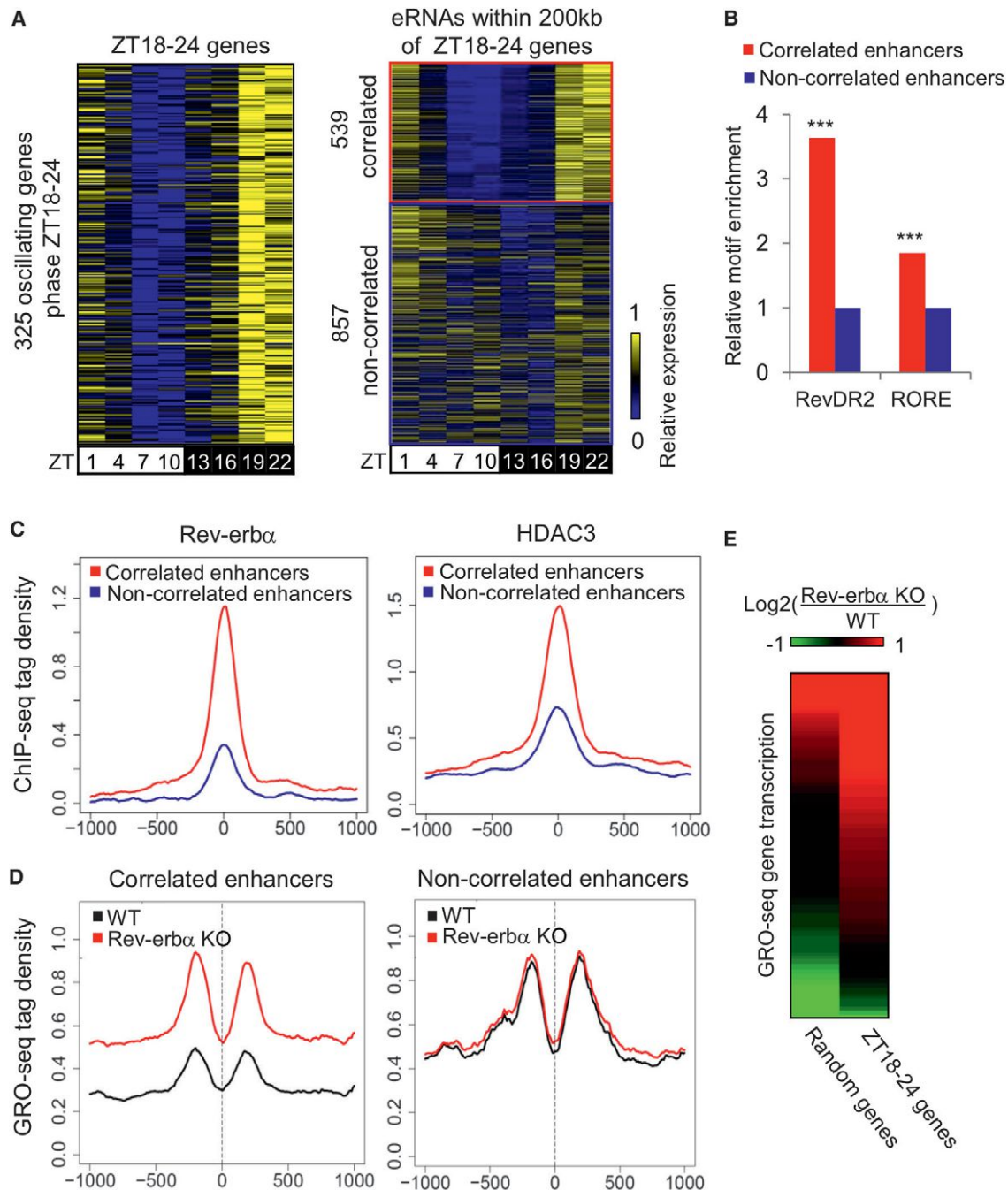


Figure 4. Phase Correlation between eRNA and Gene Body Transcription Marks Functional Enhancers of Circadian Genes

(A) Heatmap of the relative transcription of 325 circadian genes in phase ZT18–ZT24 (left) and their neighboring eRNAs (right). A total of 539 eRNAs in correlated phase are shown in the red box while 857 noncorrelated eRNAs are in the blue box.

(B) Enrichment of RevDR2 and RORE motif in correlated eRNA loci relative to noncorrelated eRNA loci (hypergeometric test, ***p < 0.001).

(C) ChIP-seq tag density of Rev-erbα (left) and HDAC3 (right) in 2 kb windows surrounding correlated (red) and noncorrelated eRNA loci (blue). Y axis shows the average tag count per 10 bp bin normalized to 10 million total reads.

(D) Comparison of GRO-seq tag density (RPTM per 10 bp bin in 2 kb window) surrounding correlated (left) and noncorrelated (right) eRNA loci in WT and Rev-erbα^{-/-} livers at ZT10.

(E) Heatmap of transcriptional changes between WT and Rev-erbα^{-/-} livers at ZT10, for the 325 circadian genes in phase ZT18–ZT24 (right column), compared to the same number of random genes (left column). Data are expressed as log₂ fold change.

See also Figure S4.

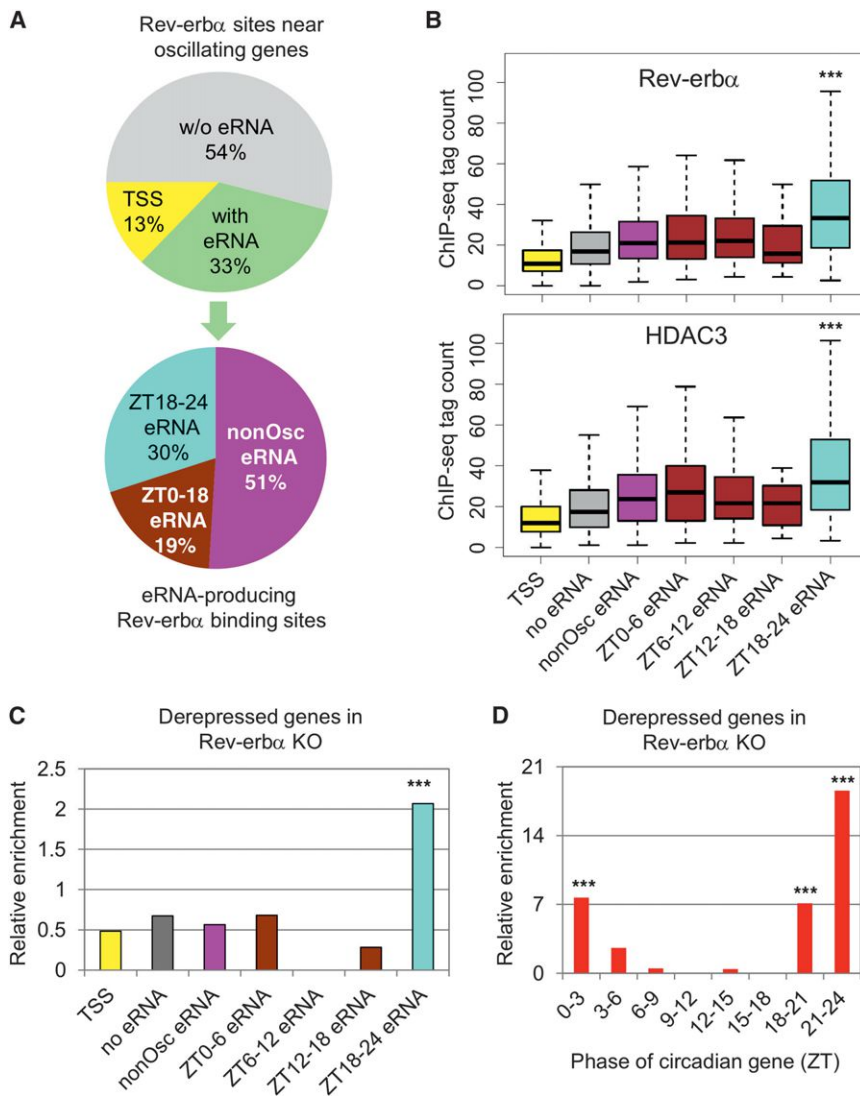


Figure 5. Circadian eRNAs Reveal the Function of the Rev-erb α Cistrome at Oscillating Genes

(A) Distribution of Rev-erb α ChIP-seq peaks near circadian genes (upper panel) and subdistribution of eRNA-producing Rev-erb α peaks near circadian genes (lower panel).

(B) Boxplot showing Rev-erb α and HDAC3 peak height at binding sites from (A). Y axis indicates normalized tag count in each peak (RPTM) (** $p < 0.001$, one-way ANOVA and Tukey's test).

(C) Enrichment of derepressed genes in Rev-erb α ^{-/-} mice at circadian genes bound by different Rev-erb α peaks from (A) relative to a random set of Rev-erb α peaks (hypergeometric test, ** $p < 0.001$).

(D) Enrichment of derepressed genes in Rev-erb α ^{-/-} mice in 8 groups of circadian genes with indicated phases relative to randomly selected genes (hypergeometric test, ** $p < 0.001$). See also Figure S5 and Table S4.

To identify putative functional E4BP4 sites, we analyzed the complete set of E4BP4 ChIP-seq peaks for those with higher eRNA levels at ZT9–ZT15 (ZT10/ZT22 > 3 or ZT13/ZT1 > 3). These sites, which we refer to as “E4BP4+eRNA” sites, were enriched 2-fold around genes downregulated in Rev-erb α ^{-/-} mice (Figure 6C), demonstrating a significant association between E4BP4 binding and gene regulation downstream of Rev-erb α . Transcriptome profiles from livers of WT mice (Hughes et al., 2009) confirmed that putative E4BP4 target genes (downregulated in Rev-erb α ^{-/-} livers and near E4BP4+eRNA sites) were generally circadian with average peak and trough expression in phase with

(Tables S4A–S4C). Profiling of eRNAs near genes that were downregulated in the Rev-erb α ^{-/-} livers revealed a marked and specific enrichment for phases between ZT9 and ZT15 (Figure 6A), which were shown earlier to be enriched for the D-box motif and binding of the D-box repressor E4BP4 (Figure 3C).

We hypothesized that, by controlling the circadian expression of E4BP4, Rev-erb α indirectly dictated the circadian expression of a large set of genes controlled by D-box enhancers whose expression would thus be in phase with Rev-erb α . Indeed, E4BP4 gene expression was circadian in WT mouse livers but constitutively elevated in Rev-erb α ^{-/-} mice (Figure 6B), consistent with a previous report (Duez et al., 2008). Furthermore, Rev-erb α bound along with its NCoR-HDAC3 corepressor complex to several sites at the *E4BP4* (*Nfil3*) locus, suggesting that E4BP4 expression is directly controlled by Rev-erb α (Figure S6A). By contrast, there were weaker changes in hepatic expression of D-box activating factors *Dbp*, *Tef*, and *Hlf* in livers of Rev-erb α ^{-/-} mice, and the expression of these factors remained circadian with similar phases (Figure S6B).

Rev-erb α and E4BP4 levels, respectively (Figure 6D, green line). The average GRO-seq transcription profile for this same group of genes showed a similar pattern over a 24 hr cycle (Figure 6D, blue line). Both patterns are consistent with direct repression by E4BP4 leading to circadian oscillation in phase with Rev-erb α protein levels. In contrast, Rev-erb α target genes (upregulated in Rev-erb α ^{-/-} livers and near Rev-erb α sites overlapping ZT18–ZT24 eRNAs) were on average antiphase to Rev-erb α expression in WT livers, consistent with direct transcriptional repression by Rev-erb α (Figure 6E). As a control, genes that were expressed near oscillating eRNAs, but unchanged in the Rev-erb α ^{-/-} livers, were not systematically phased relative to Rev-erb α or E4BP4 levels (Figure 6F).

These findings support a model in which Rev-erb α indirectly activates genes in phase ZT9–ZT15 by repressing the D-box repressor E4BP4. Such a model predicts that E4BP4 target genes would be constitutively downregulated in Rev-erb α ^{-/-} livers, with increased E4BP4 binding at nearby functional sites. Indeed, expression profiling over a 24 hr cycle revealed that

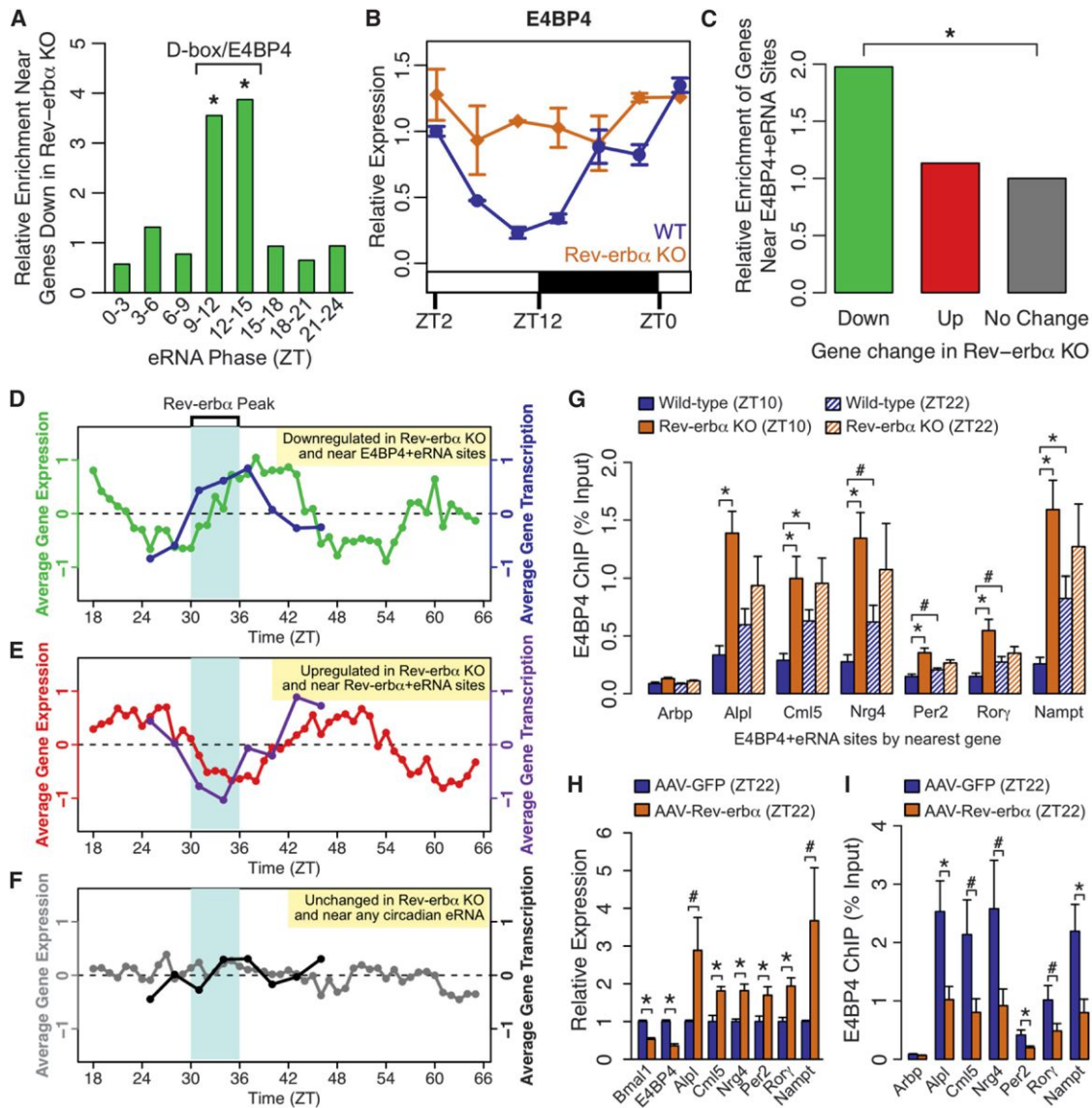


Figure 6. E4BP4 Functions Downstream of Rev-erb α

(A) Enrichment of oscillating eRNAs in each phase group near genes downregulated in Rev-erb α ^{-/-} livers relative to control genes. Significantly enriched phases are noted as corresponding to D-box/E4BP4-enriched phase group. (hypergeometric test, **p* < 0.05).

(B) mRNA expression of E4BP4/Nfi3 in WT and Rev-erb α ^{-/-} livers measured by RT-qPCR throughout the day. Data are expressed as mean \pm SEM (n = 2 per time point and genotype) normalized to the first WT time point.

(C) Enrichment of E4BP4+eRNA bound genes among those downregulated (green) or upregulated (red) in Rev-erb α ^{-/-} livers relative to unchanged genes (gray) (hypergeometric test, **p* < 0.05).

(D–F) Average circadian expression profiles in WT mouse livers (Hughes et al., 2009) and corresponding transcription profiles by GRO-seq for (D) genes downregulated in Rev-erb α ^{-/-} livers within 200 kb of E4BP4 binding at ZT9–ZT15 circadian eRNAs, (E) genes upregulated in Rev-erb α ^{-/-} livers within 200 kb of Rev-erb α binding at ZT18–ZT24 circadian eRNAs, and (F) nonregulated control genes (expressed in liver within 200 kb of circadian eRNA in any phase).

(G) ChIP-qPCR of E4BP4 binding at genes downregulated in Rev-erb α ^{-/-} livers at ZT10. Binding is shown at ZT10 (solid bars) and ZT22 (hatched bars) in WT (blue) and Rev-erb α ^{-/-} (orange) livers. Data are expressed as mean \pm SEM (one-way ANOVA, **p* < 0.05, #*p* < 0.1, n = 3–4 per group).

(H) mRNA expression measured by RT-qPCR in liver overexpressing Rev-erb α (mice injected with AAV-Tbg-Rev-erb α) or control liver (mice injected with AAV-Tbg-GFP) at ZT22. Data are expressed as mean \pm SEM (one-way ANOVA, **p* < 0.05, #*p* < 0.1; n = 6 per group).

(I) ChIP-qPCR of E4BP4 binding at same sites as (G) in liver overexpressing Rev-erb α (orange) or control liver (blue). Data are expressed as mean \pm SEM (one-way ANOVA, **p* < 0.05, #*p* < 0.1, n = 5–6 per group).

See also Figure S6.

genes near E4BP4+eRNA sites showed attenuated rhythmic expression in *Rev-erb α ^{-/-}* livers (Figure S6C). Furthermore, E4BP4 genomic binding was increased at ZT10 and no longer circadian at these sites in *Rev-erb α ^{-/-}* livers (Figure 6G).

We also tested the effect of ectopic expression of *Rev-erb α* in mouse livers on E4BP4 expression and function. Interrogation of data from a previously published experiment (Kornmann et al., 2007) revealed upregulation of the genes putatively controlled by E4BP4 in livers constitutively expressing *Rev-erb α* , particularly at the physiological peak time of E4BP4 expression (Figure S6D). Indeed, while constitutive expression of *Rev-erb α* in mouse liver repressed its direct targets such as *Bmal1* and *E4BP4/Nfil3*, it upregulated E4BP4 target genes at ZT22 (Figure 6H). This effect was much less apparent at ZT10 when E4BP4 is already at physiologically low levels (Figure S6E). Importantly, E4BP4 binding at putative functional sites near these genes was reduced at ZT22, consistent with loss of repression by E4BP4 at the implicated D-box elements (Figure 6I). These results strongly suggest that E4BP4 functions downstream of *Rev-erb α* , via sites transcribing eRNA in phase ZT9–ZT15, to repress the genes that are downregulated in *Rev-erb α ^{-/-}* livers and upregulated when *Rev-erb α* is overexpressed.

Circadian eRNAs Define Functional Cistromes that Distinguish CLOCK and *Rev-erb α* Target Genes

CLOCK and *Rev-erb α* have opposite effects on gene transcription; however, their maximal binding to the genome occur in roughly the same time window (ZT8–ZT10) (Cho et al., 2012; Feng et al., 2011; Koike et al., 2012). ChIP-seq results suggest that 80% of genes bound by CLOCK within 200 kb of TSS were also bound by *Rev-erb α* (Figure S7A), resulting in 15%–35% of circadian genes in different phases cobound by these two factors (Figure S7B). The question as to how co-occurrence of CLOCK and *Rev-erb α* binding affects rhythmic gene transcription remains unsolved (Zhao et al., 2014).

Having demonstrated that functional *Rev-erb α* sites marked by ZT18–ZT24 eRNAs correlated with target gene phase (Figures 4 and 5), we tested whether eRNAs oscillating in other phases could identify the functional cistromes of other clock components. To this end, we analyzed published microarray data measuring gene expression in livers of WT and *Clock* mutant mice (Miller et al., 2007). We first noted that genes downregulated in the *Clock* mutant mice were significantly enriched for circadian eRNAs in the phase ZT6–ZT9 compared to control genes (Figure S7C), corresponding to the enrichment of E-box motif and CLOCK binding. We then selected putatively functional CLOCK sites (Koike et al., 2012) producing eRNAs in phase with CLOCK binding (Table S5, eRNA level ZT7/ZT19 > 3 or ZT10/ZT22 > 3) and correlated with nearby gene transcription and compared these sites to the remainder of the CLOCK cistrome.

Target genes within 200 kb of putatively functional CLOCK sites showed rhythmic mRNA expression in WT mice (Miller et al., 2007), peaking at the time point corresponding to ZT10 in our studies (Figure 7A, yellow line). These genes also showed reduced expression overall in *Clock* mutant mice, particularly at time points corresponding to ZT6 and ZT10 (Figure 7A, orange line). By comparison, genes near other CLOCK sites showed

weaker average rhythm and weaker average reduction in *Clock* mutant mice (Figure 7B). Further confirming that CLOCK sites marked by in phase eRNA represent the functional subset of the CLOCK cistrome, target genes near these sites are significantly enriched for circadian genes specifically in phases ZT6–ZT12, but not opposing phases (Figure 7C) and are also significantly enriched for genes downregulated >1.5-fold in *Clock* mutants (Figure 7D). The fact that mRNA levels of some CLOCK target genes cycle in phases ZT9–ZT12 is likely due to delays in the phase of mature mRNA oscillations relative to nascent transcription, as noted in previous studies (Menet et al., 2012). Taken together, these results demonstrate that CLOCK sites marked by in phase eRNAs represent the functional component of the total cistrome.

To examine whether CLOCK and *Rev-erb α* are both functional at cobound circadian genes, functional binding sites of each factor were mapped to their closest circadian genes. CLOCK binding sites at TSS were included in this analysis as they are also enriched at genes downregulated in *Clock* mutant mouse livers (Figure S7D), consistent with previous studies (Rey et al., 2011). Remarkably, the majority of cobound circadian genes contained functional binding sites of only one factor but not both, with genes around phase ZT6–ZT9 and ZT18–ZT24 most enriched for functional CLOCK and *Rev-erb α* sites, respectively (Figure 7E). These findings suggest exclusive functions of either CLOCK or *Rev-erb α* at most cobound genes. Consistent with this notion, expression profiling showed that cobound genes exclusively carrying functional CLOCK sites, such as *Nr1d1*, *Nr1d2*, and *Tef*, are deactivated in *Clock* mutant mice, while those only carrying functional *Rev-erb α* sites, such as *Cry1* and *E4BP4*, are derepressed in *Rev-erb α ^{-/-}* mice (Figures 7F and S7E). Therefore, despite frequent colocalization of their binding, CLOCK and *Rev-erb α* control distinct sets of circadian genes that can be predicted from their regulation of eRNAs.

DISCUSSION

Unbiased analysis of the nascent transcription of over 5,000 circadian eRNAs and the TF motifs at these sites has allowed us to identify the direct genomic targets of multiple circadian regulators in mouse liver. Circadian eRNA loci are enriched for enhancer marks, the phase of eRNA oscillation correlated with that of nearby genes, and knockout studies demonstrated the causal relationship between TF binding and the transcriptional regulation at enhancers and the genes they control. These results informed the comparison of cistromes with gene expression and thus revealed the functional cistromes of multiple TFs that bind at thousands of genomic sites in liver.

Previous genomic studies of circadian gene regulation have focused primarily on the core clock components BMAL1/CLOCK, which bind DNA with a uniform genome-wide phase peaking at ZT6–ZT9 (Hatanaka et al., 2010; Koike et al., 2012; Menet et al., 2012; Rey et al., 2011; Yoshitane et al., 2014), yet only a small fraction of circadian gene transcription is in this phase. Our data suggest that only the genes with phase ZT6–ZT9 are the true BMAL1/CLOCK targets, while many other genes are bound, but not controlled, by BMAL1/CLOCK possibly due to inactive binding or long distance looping to different genes.

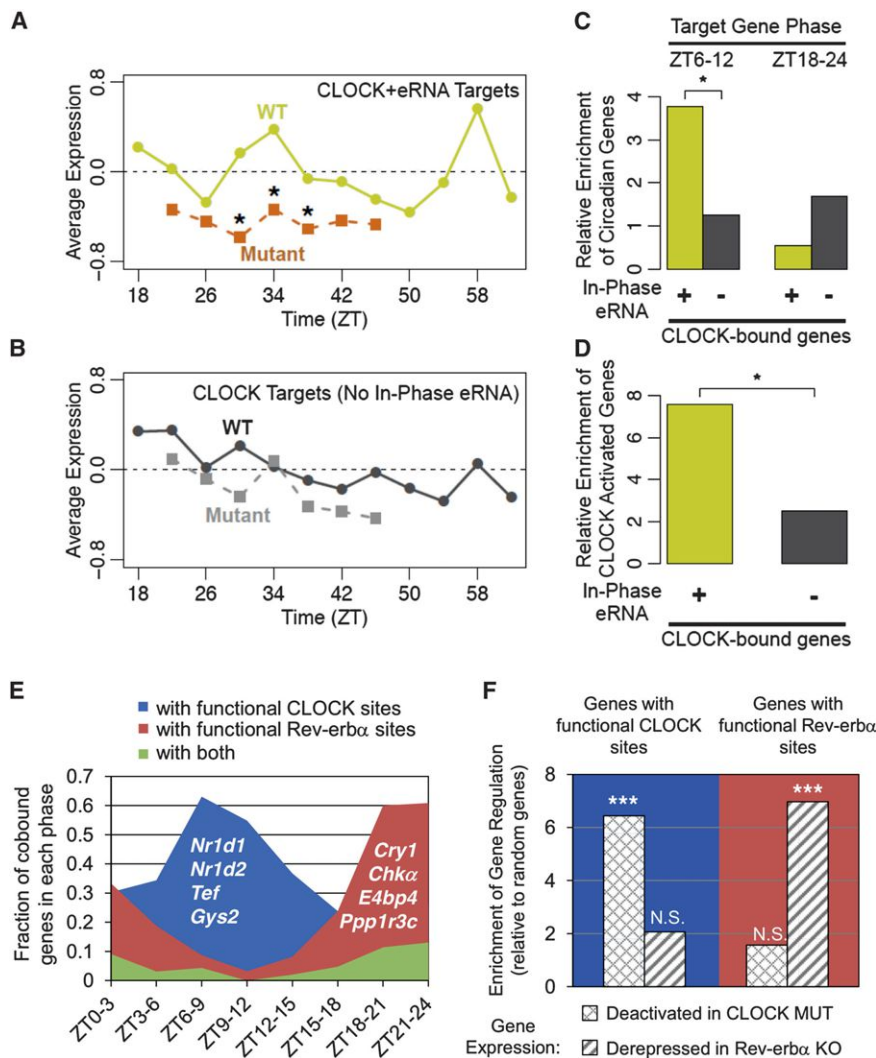


Figure 7. Circadian eRNAs Define Functional Cistromes that Distinguish CLOCK and Rev-erb α Target Genes

(A) Average expression of genes within 200 kb of CLOCK binding sites producing eRNA in phase with CLOCK binding and target gene expression in WT (yellow line) and *Clock* mutant (orange line) mouse livers from Miller et al. (2007) (Wilcoxon test of gene fold-change distribution versus matching time points in (B), * $p < 0.05$).

(B) Average expression of genes within 200 kb of CLOCK binding sites lacking in-phase eRNA in WT (dark gray line) and *Clock* mutant (light gray line) mouse livers from Miller et al. (2007).

(C) Enrichment of circadian genes expressed in phase with CLOCK binding (ZT6–ZT12) or anti-phase to CLOCK binding (ZT18–ZT24) for the gene groups used in (A) (yellow) and (B) (gray) relative to random genes (hypergeometric test, * $p < 0.05$).

(D) Enrichment of genes downregulated in *Clock* mutant livers among the gene groups used in (A–C) relative to random genes (hypergeometric test, * $p < 0.05$).

(E) Fraction of oscillating genes cobound by CLOCK and Rev-erb α that are within 200 kb of TF binding sites producing rhythmic eRNA in phase with CLOCK activation (blue), Rev-erb α repression (red), or both (green). Oscillating genes are divided according to their phases. Representative genes are noted in each group.

(F) Enrichment of CLOCK and Rev-erb α regulated genes (expression fold change in mutant >95% of random genes) in those with eRNA predicted functional binding sites in (E), relative to random genes (hypergeometric test, *** $p < 0.001$, not significant [N.S.] $p > 0.05$).

See also Figure S7 and Table S5.

Moreover, despite extensive binding region overlap with Rev-erb α (Cho et al., 2012), whose repressive activity would conflict with activation by BMAL1/CLOCK, our results demonstrate on a genome-wide scale that enhancer activity is primarily controlled by one factor or the other.

Importantly, our unbiased identification of enhancers revealed not only the ZT6–ZT9 enhancers marked by E-box motifs and bound by BMAL1, NPAS2, and CLOCK but also more abundant sets of enhancers in other phases. Those peaking at ZT0–ZT3, ZT9–ZT15, and ZT18–ZT24 were enriched for ETS, D-box, and RevDR2/RORE motifs, respectively. The ETS motif is recognized by a large family of TFs (Hollenhorst et al., 2011), some of which have recently been implicated in circadian biology and will be the focus of future research (Anafi et al., 2014; Ciarleglio et al., 2014). Moreover, by integration of enhancer sites with cistromic data, E4BP4 emerged as a key regulator of the ZT9–ZT15 D-box enhancers in normal liver, as well in the *Rev-erb α* ^{−/−} livers, and Rev-erb α was clearly a strong antiphase repressor bound to RevDR2/RORE sites at ZT18–ZT24 enhancers.

Interestingly, the phase of circadian enhancers exhibited an uneven distribution, with 42% of circadian eRNAs peaking during the late night (ZT18–ZT24), while rhythmic gene transcription was more evenly distributed across all phases. A possible explanation is that the regulation of genes whose transcription peaks in the light cycle might be primarily regulated at promoters. For example, BMAL1 controls gene transcription at both promoters and enhancers (Rey et al., 2011), whereas Rev-erb α , the main controller of the ZT18–ZT24 phase, binds mainly intergenically (Feng et al., 2011; Lam et al., 2013). The overabundance of enhancers in phase ZT18–ZT24 is surprising, yet remarkably consistent with the previously unexplained finding of Koike et al. (2012) that the global peak of initiated Pol2 occurs at ~ZT22–ZT24.

Analysis of oscillating eRNAs in mice fed normal chow ad libitum did not reveal the motifs for TFs previously suggested to entrain liver circadian gene expression to feeding/fasting cycles, such as CREB, SREBP, PPARs, and FOXO1 (Adamovich et al., 2014; Eckel-Mahan et al., 2013; Vollmers et al., 2009). Some of these TFs, such as CREB and SREBP, bind preferentially to promoters of target genes (Everett et al., 2013; Gilardi et al., 2014; Seo et al., 2009), which would not be captured by analysis of

eRNAs. Phase-specific enrichment could also have been masked by motifs bound by constitutive liver TFs, such as HNF4A and FOXA1, that bind at enhancers in all phases. It will be interesting to profile eRNAs under altered dietary conditions in future studies to examine the interplay between metabolic cues and circadian rhythms at enhancers.

Rev-erb α expression and repressive function peaks at ZT10 in liver, thereby orchestrating circadian transcription in the opposing phase (ZT22) (Feng et al., 2011). Consistent with this, recruitment of Rev-erb α and its corepressor was strongest at sites of ZT18–ZT24 eRNA transcription. It should be noted that the entire Rev-erb α cistrome in liver includes thousands of other binding sites, with <10% characterized by rhythmic eRNAs anti-phase to Rev-erb α binding. Deletion of Rev-erb α specifically activated transcription of these eRNAs, as well as the genes they control, thus clearly delineating the functional component of the Rev-erb α cistrome.

In addition to the direct regulation of circadian genes antiphase to Rev-erb α expression, we uncovered a large set of in-phase circadian transcripts that were downregulated in the absence of Rev-erb α , contrary to its powerful repressive function. Functional enhancer analysis suggested that the downregulated genes in Rev-erb α ^{-/-} mice were mediated by D-box factors, including E4BP4, a direct target of Rev-erb α . While the direct regulation of E4BP4 by Rev-erb α has been recognized (Duez et al., 2008), a relatively small number of E4BP4 target genes have been identified in liver, based primarily on in vitro studies of proximal promoter constructs (Tong et al., 2010; Ueda et al., 2005). Our study includes a ChIP-seq study of E4BP4 in liver, and our integrative analysis demonstrates the extensive, genome-wide effects of this pathway, revealing how a single TF, such as Rev-erb α , can regulate opposing phases of circadian gene expression by its direct and indirect actions.

Together, the present studies reveal mechanisms for generating and coordinating multiple phases of circadian transcription in a single organ. They also demonstrate that the unbiased analysis of enhancer activity and correlated gene expression is a powerful method of discovering relevant TFs and their specific functional cistromes, which can be more generally applied to understanding the transcriptional regulation of physiology and disease states.

EXPERIMENTAL PROCEDURES

Mice

WT C57Bl/6 mice were purchased from the Jackson Laboratories. The Rev-erb α ^{-/-} mice were obtained from B. Vennström and backcrossed greater than or equal to seven generations with C57Bl/6 mice. WT and mutant male mice (10- to 12-week-old) were housed under standard 12 hr light/12 hr dark cycles, with lights on (ZT0) at 7 a.m. and lights off (ZT12) at 7 p.m. and euthanized at indicated times. All animal care and procedures followed the guidelines of the Institutional Animal Care and Use Committee of the University of Pennsylvania.

Antibodies

E4BP4 antibodies (Santa Cruz sc-9550 and sc-9549) were mixed in 1:1 ratio for ChIP. ROR α antibody was purchased from Santa Cruz (sc-6062).

GRO-Seq

The GRO-seq was performed as previously described (Core et al., 2008; Step et al., 2014; Wang et al., 2011). Raw data are available in Gene Expression

Omibus (GEO) (GSE59486). See also Extended Experimental Procedures.

De Novo Identification of eRNAs

A pipeline was constructed for genome-wide de novo identification of eRNA loci. See also Extended Experimental Procedures.

Analysis of Oscillating Gene Transcripts and eRNAs

RPKM values across all time points for each transcript and eRNA feature were analyzed for significant circadian oscillations using JTK_CYCLE (Hughes et al., 2010). Motif mining at oscillating eRNAs was performed by applying HOMER to the 500bp window centered on each locus. See also Extended Experimental Procedures.

Gene and eRNA Expression Analysis

Total RNA was extracted from liver using the RNeasy Mini Kit (QIAGEN) and treated with DNase (QIAGEN). RNA was reverse transcribed using the High-Capacity cDNA Reverse Transcription Kit (Applied Biosystems). Quantitative PCR was performed with Power SYBR Green PCR Mastermix on the PRISM 7500 (Applied Biosystems) and analyzed by the standard curve method. Gene or eRNA expression was normalized to mRNA levels of housekeeping gene *36B4* (*Arbp*). Primer sequences can be found in Table S1D.

Microarray Analysis

Microarray analysis of WT and Rev-erb α ^{-/-} livers (n = 5) was performed by the Penn Microarray Core. Raw data are available from GEO (GSE59460). See also Extended Experimental Procedures.

ChIP

ChIP-qPCR and ChIP-seq experiments were performed as described (Feng et al., 2011) with minor changes. Raw data for ROR α and E4BP4 ChIP-seq are available in GEO (GSE59486). See also Extended Experimental Procedures.

ChIP-Seq Data Analysis

Sequenced reads were aligned to the mouse reference genome (mm9) and peak calling was performed with HOMER (Heinz et al., 2010). Sources of public ChIP-seq data analyzed are listed in Table S2B. See also Extended Experimental Procedures.

Liver-Specific Gene Expression

Flag-Rev-erb α and GFP cDNAs were subcloned into hepatocyte-specific AAV vector AAV8-Tbg (Bell et al., 2011) and tail veins were injected with 1×10^{12} genome copies per mouse. Livers were harvested 2 weeks after injection.

ACCESSION NUMBERS

The GEO accession number for the GRO-seq and ChIP-seq data reported in this paper is GSE59486. The GEO accession number for the microarray data reported in this paper is GSE59460.

SUPPLEMENTAL INFORMATION

Supplemental Information includes Extended Experimental Procedures, seven figures, and five tables and can be found with this article online at <http://dx.doi.org/10.1016/j.cell.2014.10.022>.

AUTHOR CONTRIBUTIONS

B.F., L.J.E., J.J., and M.A.L. conceived the project design and wrote the manuscript. B.F. performed analysis of GRO-seq data. B.F., L.J.E., and A.R. performed integrative genomic analyses. J.J. performed GRO-seq experiments. E.B. and L.J.E. performed E4BP4 ChIP-seq. D.F. performed ROR α ChIP-seq. J.J., E.B., and Z.G.H. performed RT-qPCR and ChIP-qPCR experiments. S.M.A. and Z.S. performed AAV overexpression experiments.

ACKNOWLEDGMENTS

We thank the Functional Genomics Core (J. Schug and K. Kaestner) and Viral Vector Core (J. Johnston and A. Sandhu) of the Penn Diabetes Research Center (P30 DK19525) for next-generation sequencing and AAV production, respectively. We also thank the Penn Microarray Core for microarray analysis. We thank Dr. Ken Zaret for critical reading of the manuscript. This work was supported by NIH grants R01 DK45586 (M.A.L.), F32 DK095526 (L.J.E.), K99 DK099443 (Z.S.), and F32 DK095563 (Z.G.H.), and the JPB Foundation.

Received: April 21, 2014

Revised: August 28, 2014

Accepted: September 24, 2014

Published: November 20, 2014

REFERENCES

- Adamovich, Y., Roussou-Noori, L., Zwihaft, Z., Neufeld-Cohen, A., Golik, M., Kraut-Cohen, J., Wang, M., Han, X., and Asher, G. (2014). Circadian clocks and feeding time regulate the oscillations and levels of hepatic triglycerides. *Cell Metab.* 19, 319–330.
- Anafi, R.C., Lee, Y., Sato, T.K., Venkataraman, A., Ramanathan, C., Kavakli, I.H., Hughes, M.E., Baggs, J.E., Growe, J., Liu, A.C., et al. (2014). Machine learning helps identify CHRONO as a circadian clock component. *PLoS Biol.* 12, e1001840.
- Asher, G., and Schibler, U. (2011). Crosstalk between components of circadian and metabolic cycles in mammals. *Cell Metab.* 13, 125–137.
- Bell, P., Gao, G., Haskins, M.E., Wang, L., Sleeper, M., Wang, H., Calcedo, R., Vandenberghe, L.H., Chen, S.J., Weiss, C., et al. (2011). Evaluation of adeno-associated viral vectors for liver-directed gene transfer in dogs. *Hum. Gene Ther.* 22, 985–997.
- Bugge, A., Feng, D., Everett, L.J., Briggs, E.R., Mullican, S.E., Wang, F., Jager, J., and Lazar, M.A. (2012). Rev-erb α and Rev-erb β coordinately protect the circadian clock and normal metabolic function. *Genes Dev.* 26, 657–667.
- Cho, H., Zhao, X., Hatori, M., Yu, R.T., Barish, G.D., Lam, M.T., Chong, L.W., DiTacchio, L., Atkins, A.R., Glass, C.K., et al. (2012). Regulation of circadian behaviour and metabolism by REV-ERB- α and REV-ERB- β . *Nature* 485, 123–127.
- Ciarleglio, C.M., Resuehr, H.E., Axley, J.C., Deneris, E.S., and McMahon, D.G. (2014). Pet-1 deficiency alters the circadian clock and its temporal organization of behavior. *PLoS ONE* 9, e97412.
- Core, L.J., Waterfall, J.J., and Lis, J.T. (2008). Nascent RNA sequencing reveals widespread pausing and divergent initiation at human promoters. *Science* 322, 1845–1848.
- Cowell, I.G., Skinner, A., and Hurst, H.C. (1992). Transcriptional repression by a novel member of the bZIP family of transcription factors. *Mol. Cell. Biol.* 12, 3070–3077.
- Creyghton, M.P., Cheng, A.W., Welstead, G.G., Kooistra, T., Carey, B.W., Steine, E.J., Hanna, J., Lodato, M.A., Frampton, G.M., Sharp, P.A., et al. (2010). Histone H3K27ac separates active from poised enhancers and predicts developmental state. *Proc. Natl. Acad. Sci. USA* 107, 21931–21936.
- Dibner, C., Schibler, U., and Albrecht, U. (2010). The mammalian circadian timing system: organization and coordination of central and peripheral clocks. *Annu. Rev. Physiol.* 72, 517–549.
- Duez, H., van der Veen, J.N., Duhem, C., Pourcet, B., Touvier, T., Fontaine, C., Derudas, B., Haugé, E., Havinga, R., Bloks, V.W., et al. (2008). Regulation of bile acid synthesis by the nuclear receptor Rev-erb α . *Gastroenterology* 135, 689–698.
- Eckel-Mahan, K.L., Patel, V.R., de Mateo, S., Orozco-Solis, R., Ceglia, N.J., Sahar, S., Dilag-Penilla, S.A., Dyar, K.A., Baldi, P., and Sassone-Corsi, P. (2013). Reprogramming of the circadian clock by nutritional challenge. *Cell* 155, 1464–1478.
- Everett, L.J., Le Lay, J., Lukovac, S., Bernstein, D., Steger, D.J., Lazar, M.A., and Kaestner, K.H. (2013). Integrative genomic analysis of CREB defines a critical role for transcription factor networks in mediating the fed/fasted switch in liver. *BMC Genomics* 14, 337.
- Feng, D., Liu, T., Sun, Z., Bugge, A., Mullican, S.E., Alenghat, T., Liu, X.S., and Lazar, M.A. (2011). A circadian rhythm orchestrated by histone deacetylase 3 controls hepatic lipid metabolism. *Science* 331, 1315–1319.
- Gachon, F., Olela, F.F., Schaad, O., Descombes, P., and Schibler, U. (2006). The circadian PAR-domain basic leucine zipper transcription factors DBP, TEF, and HLF modulate basal and inducible xenobiotic detoxification. *Cell Metab.* 4, 25–36.
- Giguère, V., Tini, M., Flock, G., Ong, E., Evans, R.M., and Otulakowski, G. (1994). Isoform-specific amino-terminal domains dictate DNA-binding properties of ROR α , a novel family of orphan hormone nuclear receptors. *Genes Dev.* 8, 538–553.
- Gilardi, F., Migliavacca, E., Naldi, A., Baruchet, M., Canella, D., Le Martelot, G., Guex, N., and Desvergne, B.; CyclIX Consortium (2014). Genome-wide analysis of SREBP1 activity around the clock reveals its combined dependency on nutrient and circadian signals. *PLoS Genet.* 10, e1004155.
- Hah, N., Murakami, S., Nagari, A., Danko, C.G., and Kraus, W.L. (2013). Enhancer transcripts mark active estrogen receptor binding sites. *Genome Res.* 23, 1210–1223.
- Harding, H.P., and Lazar, M.A. (1995). The monomer-binding orphan receptor Rev-Erb represses transcription as a dimer on a novel direct repeat. *Mol. Cell. Biol.* 15, 4791–4802.
- Hatanaka, F., Matsubara, C., Myung, J., Yoritaka, T., Kamimura, N., Tsutsumi, S., Kanai, A., Suzuki, Y., Sassone-Corsi, P., Aburatani, H., et al. (2010). Genome-wide profiling of the core clock protein BMAL1 targets reveals a strict relationship with metabolism. *Mol. Cell. Biol.* 30, 5636–5648.
- Heinz, S., Benner, C., Spann, N., Bertolino, E., Lin, Y.C., Laslo, P., Cheng, J.X., Murre, C., Singh, H., and Glass, C.K. (2010). Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Mol. Cell* 38, 576–589.
- Hollenhorst, P.C., McIntosh, L.P., and Graves, B.J. (2011). Genomic and biochemical insights into the specificity of ETS transcription factors. *Annu. Rev. Biochem.* 80, 437–471.
- Hughes, M.E., DiTacchio, L., Hayes, K.R., Vollmers, C., Pulivarthy, S., Baggs, J.E., Panda, S., and Hogenesch, J.B. (2009). Harmonics of circadian gene transcription in mammals. *PLoS Genet.* 5, e1000442.
- Hughes, M.E., Hogenesch, J.B., and Kornacker, K. (2010). JTK_CYCLE: an efficient nonparametric algorithm for detecting rhythmic components in genome-scale data sets. *J. Biol. Rhythms* 25, 372–380.
- Kim, T.K., Hemberg, M., Gray, J.M., Costa, A.M., Bear, D.M., Wu, J., Harmin, D.A., Laptewicz, M., Barbara-Haley, K., Kuersten, S., et al. (2010). Widespread transcription at neuronal activity-regulated enhancers. *Nature* 465, 182–187.
- Koike, N., Yoo, S.H., Huang, H.C., Kumar, V., Lee, C., Kim, T.K., and Takahashi, J.S. (2012). Transcriptional architecture and chromatin landscape of the core circadian clock in mammals. *Science* 338, 349–354.
- Kornmann, B., Schaad, O., Bujard, H., Takahashi, J.S., and Schibler, U. (2007). System-driven and oscillator-dependent circadian transcription in mice with a conditionally active liver clock. *PLoS Biol.* 5, e34.
- Lam, M.T., Cho, H., Lesch, H.P., Gosselin, D., Heinz, S., Tanaka-Oishi, Y., Benner, C., Kaikkonen, M.U., Kim, A.S., Kosaka, M., et al. (2013). Rev-Erbs repress macrophage gene expression by inhibiting enhancer-directed transcription. *Nature* 498, 511–515.
- Li, S., and Hunger, S.P. (2001). The DBP transcriptional activation domain is highly homologous to that of HLF and TEF and is not responsible for the tissue type-specific transcriptional activity of DBP. *Gene* 263, 239–245.
- Li, W., Notani, D., Ma, Q., Tanasa, B., Nunez, E., Chen, A.Y., Merkurjev, D., Zhang, J., Ohgi, K., Song, X., et al. (2013). Functional roles of enhancer RNAs for oestrogen-dependent transcriptional activation. *Nature* 498, 516–520.
- Menet, J.S., Rodriguez, J., Abruzzi, K.C., and Rosbash, M. (2012). Nascent-Seq reveals novel features of mouse circadian transcriptional regulation. *eLife* 1, e00011.

- Miller, B.H., McDearmon, E.L., Panda, S., Hayes, K.R., Zhang, J., Andrews, J.L., Antoch, M.P., Walker, J.R., Esser, K.A., Hogenesch, J.B., and Takahashi, J.S. (2007). Circadian and CLOCK-controlled regulation of the mouse transcriptome and cell proliferation. *Proc. Natl. Acad. Sci. USA* **104**, 3342–3347.
- Mitsui, S., Yamaguchi, S., Matsuo, T., Ishida, Y., and Okamura, H. (2001). Antagonistic role of E4BP4 and PAR proteins in the circadian oscillatory mechanism. *Genes Dev.* **15**, 995–1006.
- Panda, S., Antoch, M.P., Miller, B.H., Su, A.I., Schook, A.B., Straume, M., Schultz, P.G., Kay, S.A., Takahashi, J.S., and Hogenesch, J.B. (2002). Coordinated transcription of key pathways in the mouse by the circadian clock. *Cell* **109**, 307–320.
- Peek, C.B., Ramsey, K.M., Marcheva, B., and Bass, J. (2012). Nutrient sensing and the circadian clock. *Trends Endocrinol. Metab.* **23**, 312–318.
- Reppert, S.M., and Weaver, D.R. (2001). Molecular analysis of mammalian circadian rhythms. *Annu. Rev. Physiol.* **63**, 647–676.
- Rey, G., Cesbron, F., Rougemont, J., Reinke, H., Brunner, M., and Naef, F. (2011). Genome-wide and phase-specific DNA-binding rhythms of BMAL1 control circadian output functions in mouse liver. *PLoS Biol.* **9**, e1000595.
- Ripperger, J.A., and Schibler, U. (2001). Circadian regulation of gene expression in animals. *Curr. Opin. Cell Biol.* **13**, 357–362.
- Ripperger, J.A., and Schibler, U. (2006). Rhythmic CLOCK-BMAL1 binding to multiple E-box motifs drives circadian Dbp transcription and chromatin transitions. *Nat. Genet.* **38**, 369–374.
- Seo, Y.K., Chong, H.K., Infante, A.M., Im, S.S., Xie, X., and Osborne, T.F. (2009). Genome-wide analysis of SREBP-1 binding in mouse liver chromatin reveals a preference for promoter proximal binding to a new motif. *Proc. Natl. Acad. Sci. USA* **106**, 13765–13769.
- Step, S.E., Lim, H.W., Marinis, J.M., Prokesch, A., Steger, D.J., You, S.H., Won, K.J., and Lazar, M.A. (2014). Anti-diabetic rosiglitazone remodels the adipocyte transcriptome by redistributing transcription to PPAR γ -driven enhancers. *Genes Dev.* **28**, 1018–1028.
- Takahashi, J.S., Hong, H.K., Ko, C.H., and McDearmon, E.L. (2008). The genetics of mammalian circadian order and disorder: implications for physiology and disease. *Nat. Rev. Genet.* **9**, 764–775.
- Tong, X., Muchnik, M., Chen, Z., Patel, M., Wu, N., Joshi, S., Rui, L., Lazar, M.A., and Yin, L. (2010). Transcriptional repressor E4-binding protein 4 (E4BP4) regulates metabolic hormone fibroblast growth factor 21 (FGF21) during circadian cycles and feeding. *J. Biol. Chem.* **285**, 36401–36409.
- Ueda, H.R., Hayashi, S., Chen, W., Sano, M., Machida, M., Shigeyoshi, Y., Iino, M., and Hashimoto, S. (2005). System-level identification of transcriptional circuits underlying mammalian circadian clocks. *Nat. Genet.* **37**, 187–192.
- Ukai-Tadenuma, M., Kasukawa, T., and Ueda, H.R. (2008). Proof-by-synthesis of the transcriptional logic of mammalian circadian clocks. *Nat. Cell Biol.* **10**, 1154–1163.
- Vollmers, C., Gill, S., DiTacchio, L., Pulivarthy, S.R., Le, H.D., and Panda, S. (2009). Time of feeding and the intrinsic circadian clock drive rhythms in hepatic gene expression. *Proc. Natl. Acad. Sci. USA* **106**, 21453–21458.
- Wang, D., Garcia-Bassets, I., Benner, C., Li, W., Su, X., Zhou, Y., Qiu, J., Liu, W., Kaikkonen, M.U., Ohgi, K.A., et al. (2011). Reprogramming transcription by distinct classes of enhancers functionally defined by eRNA. *Nature* **474**, 390–394.
- Yoshitane, H., Ozaki, H., Terajima, H., Du, N.H., Suzuki, Y., Fujimori, T., Kosaka, N., Shimba, S., Sugano, S., Takagi, T., et al. (2014). CLOCK-controlled polyphonic regulation of circadian rhythms through canonical and noncanonical E-boxes. *Mol. Cell. Biol.* **34**, 1776–1787.
- Zhao, Q., Khorasanizadeh, S., Miyoshi, Y., Lazar, M.A., and Rastinejad, F. (1998). Structural elements of an orphan nuclear receptor-DNA complex. *Mol. Cell* **1**, 849–861.
- Zhao, X., Cho, H., Yu, R.T., Atkins, A.R., Downes, M., and Evans, R.M. (2014). Nuclear receptors rock around the clock. *EMBO Rep.* **15**, 518–528.

A Variety of Dicer Substrates in Human and *C. elegans*

Agnieszka Rybak-Wolf,^{1,3} Marvin Jens,^{1,3} Yasuhiro Murakawa,^{2,3} Margareta Herzog,¹ Markus Landthaler,^{2,*} and Nikolaus Rajewsky^{1,*}

¹Laboratory for Systems Biology of Gene Regulatory Elements

²Laboratory for RNA Biology and Posttranscriptional Regulation

Berlin Institute for Medical Systems Biology, Max-Delbrück-Center for Molecular Medicine, Robert-Rössle-Strasse 10, 13125 Berlin, Germany

³Co-first author

*Correspondence: markus.landthaler@mdc-berlin.de (M.L.), rajewsky@mdc-berlin.de (N.R.)

<http://dx.doi.org/10.1016/j.cell.2014.10.040>

SUMMARY

The endoribonuclease Dicer is known for its central role in the biogenesis of eukaryotic small RNAs/miRNAs. Despite its importance, Dicer target transcripts have not been directly mapped. Here, we apply biochemical methods to human cells and *C. elegans* and identify thousands of Dicer-binding sites. We find known and hundreds of additional miRNAs with high sensitivity and specificity. We also report structural RNAs, promoter RNAs, and mitochondrial transcripts as Dicer targets. Interestingly, most Dicer-binding sites reside on mRNAs/lncRNAs and are not significantly processed into small RNAs. These passive sites typically harbor small, Dicer-bound hairpins within intact transcripts and generally stabilize target expression. We show that passive sites can sequester Dicer and reduce microRNA expression. mRNAs with passive sites were in human and worm significantly associated with processing-body/granule function. Together, we provide the first transcriptome-wide map of Dicer targets and suggest conserved binding modes and functions outside of the miRNA pathway.

INTRODUCTION

Genes are subject to posttranscriptional regulation by small RNAs (sRNA) and RNA-binding proteins (RBPs). Over the past years, many regulatory sRNAs have been discovered (Bartel, 2009; Kim et al., 2009). Most miRNAs are generated from primary transcripts that undergo two distinct steps of processing. First, DROSHA and its partner DGCR8 release ~70 nt stem-loop precursor miRNAs (pre-miRNAs) inside of the nucleus (Lee et al., 2003). Alternatively, “mirtrons” are derived from introns in a splicing-machinery-dependent and Drosha-independent fashion (Okamura et al., 2007; Ruby et al., 2007). The pre-miRNA hairpins are further cut by Dicer (Hutvagner et al., 2001). The physical distance between the Dicer PAZ and 5′ pocket and the RNase III domains functions as a molecular “ruler” to control the product size

(Lau et al., 2012; Park et al., 2011). The resulting ~22 nt miRNA/miRNA* duplexes have a 2 nt overhang at the 3′ ends, which is a characteristic signature of Dicer processing (Bartel, 2009). The duplexes are subsequently handed over to Argonaute (AGO) proteins, and one of the two strands is selected as the mature miRNA to form active RNA-induced silencing complex (RISC), which represses the expression of target genes (Bartel, 2009).

Dicer binding sites are typically indirectly inferred by mapping sRNA sequencing data to transcripts. Two adjacent read stacks suggest a pre-miRNA that was cleaved by Dicer. A more variable 3′ than 5′ end is also typical for miRNAs (Chiang et al., 2010). By computationally scoring these and other hallmarks of Dicer processing, Dicer targets are inferred (Friedländer et al., 2012). These methods invariably make assumptions about how Dicer binds and processes substrates.

sRNA sequencing revealed diverse sRNA classes, including miRNAs that are processed from tRNAs and snoRNAs (Castellano and Stebbing, 2013; Ender et al., 2008; Friedländer et al., 2014). Currently, more than 1,800 pre-miRNAs in human and 200 in *C. elegans* are listed in miRBase (Kozomara and Griffiths-Jones, 2014). However, it remains challenging to distinguish miRNAs from fragments of other transcripts based on sRNA-sequencing data alone (Kozomara and Griffiths-Jones, 2014). There are also other types of sRNAs such as endogenous short interfering RNAs (endo-siRNAs) originating from long double-stranded RNAs (dsRNAs) generated by transposable elements or convergent transcription units (Kim et al., 2009). Also sRNAs from transcript starts and ends have been described (Seila et al., 2008; Valen et al., 2011; Zamudio et al., 2014). Recent findings highlight that primary sequence, RNA structure, size, and position of stem and loop, the accessibility of 3′/5′ ends, and cofactors are all relevant for Dicer substrate recognition and/or cleavage (Feng et al., 2012; Fukunaga et al., 2012; Gu et al., 2012; Lau et al., 2012; Park et al., 2011; Tian et al., 2014). Furthermore, in vitro studies (Feng et al., 2012) and electron microscopy (Taylor et al., 2013) show that affinity and cleavage efficiency can be uncoupled properties of the bound substrate. In *C. elegans*, Dicer binds the lncRNA *mcs-1* in vitro but does not process it to sRNA, leading to sequestration of Dicer and inhibition of its function (Hellwig and Bass, 2008).

Other reports likewise point to functions of Dicer that are not connected to sRNA production. In human cells, Dicer is present

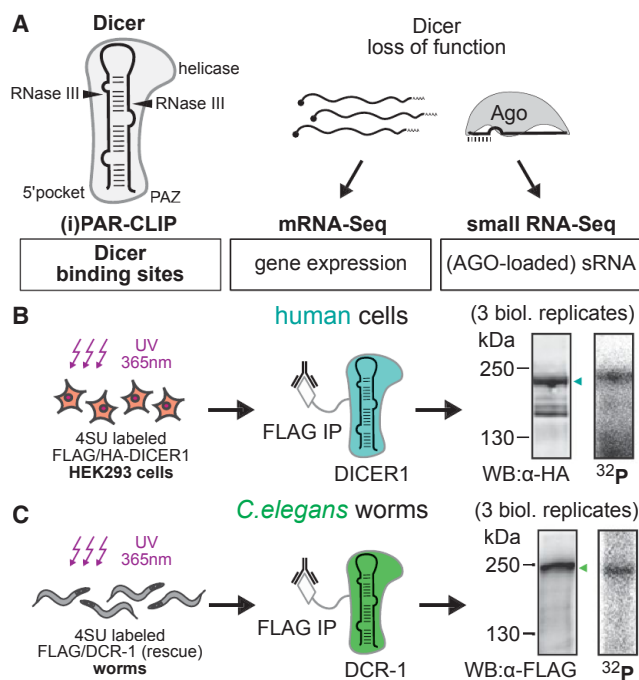


Figure 1. PAR-CLIP Identifies Transcriptome-wide Binding of Dicer
 (A) Outline of experimental setup. PAR-CLIP/iPAR-CLIP of Dicer was performed in HEK293 cells and *C. elegans*. Expressed sRNA was profiled by 5' P (C. *elegans*) or AGO2/3-IP (human) deep sequencing. The impact of Dicer and Drosha knockdowns on mRNA levels was monitored by mRNA-seq, on sRNA expression by sRNA-seq.
 (B and C) Overview of PAR-CLIP experiments in HEK293 cells (B) and in *C. elegans* (C). The right panel shows the phospho-image of SDS-PAGE-resolved, radiolabeled Dicer-RNA complex, immunoprecipitated from 4SU-labeled and crosslinked HEK293 cells (B) or worms (C). Arrows indicate the specific band of RNAs crosslinked to Dicer protein, confirmed by western blot (left).

in the nucleus (Doyle et al., 2013; Sinkkonen et al., 2010) and interacts with nuclear pore complexes (Ando et al., 2011). In *C. elegans*, Dicer localizes to germline P granules and is required for their formation (Beshore et al., 2011). Dicer is also present in chromatoid bodies of mammalian germ cells (Kotaja et al., 2006). If Dicer binding and RNA cleavage are largely uncoupled, many interactions may be undetectable with current methods.

Here, we present an in-depth identification and analysis of direct Dicer-binding sites in human embryonic kidney (HEK) 293 cells and *C. elegans* by applying PAR-CLIP/iPAR-CLIP (Hafner et al., 2010; Jungkamp et al., 2011) (Figure 1). Because mammals and nematodes both have a single full-length Dicer gene, assumed to process long dsRNAs and miRNAs, we were able to compare human and worm modes of Dicer targeting, highlighting core functions conserved during 550 million years of evolution. To measure the enzymatic activity of Dicer on its binding sites, we performed sRNA sequencing. In human cells, we deeply sequenced AGO-loaded sRNAs, enabling us to identify hundreds of additional miRNAs.

We also identified in both human and worm thousands of stably bound Dicer target sites that were not processed into sRNAs

(“passive” sites). In *C. elegans*, the top passively bound transcript was the lncRNA *mcs-1*, confirming previous in vitro data (Hellwig and Bass, 2008) and indicating that passive sites can function by sequestering Dicer. Dicer knockdown experiments in both human cells and worms suggest that passive sites, in contrast to sites that emit sRNAs, are generally stabilized by Dicer. Furthermore, we systematically compared passive Dicer targets across human and worm and discovered that messages encoding P body/granule components are statistically significantly enriched in both species.

In summary, our data reveal hundreds of additional miRNAs, non-canonical substrates for dicing, and a novel, passive mode of Dicer binding. Moreover, we identify a large number of stably bound “passive” sites inside of many mRNAs in both human and worm. We present multiple lines of evidence that these sites are bound by Dicer nearby the loop of small hairpin structures within the intact host RNA, generally stabilizing its expression. We discuss several scenarios for the function of passive sites.

RESULTS

PAR-CLIP Identifies Transcriptome-wide Binding of Dicer in Human Cells and *C. elegans*

To identify direct in vivo Dicer targets transcriptome wide, we performed PAR-CLIP (Hafner et al., 2010) in human HEK293 cells and iPAR-CLIP (Jungkamp et al., 2011) in young adult *C. elegans* worms (Figures 1B and 1C). We used HEK293 cells expressing FLAG/HA-tagged DICER1 protein (Figure S1A available online), whereas in *C. elegans*, we used a 3×FLAG/10×His-tagged DCR-1 rescue strain (Figure S1B). For HEK293 cells and worm, we combined three independent replicates each (see Figures S1C–S1F for reproducibility). After computational quality filtering, we identified ~8,500 and ~2,500 reproduced Dicer-binding sites in HEK293 cells and *C. elegans*, respectively. The crosslinking introduced specific nucleotide mutations and was not biased toward highly expressed genes (Figures S1G and S1H).

Dicer PAR-CLIP Recovers Known miRNAs

We first investigated canonical miRNAs, exemplified by the human “oncomir” cluster miR-17–92 (Figure 2A) and the miR-35–41 cluster in *C. elegans* (Figure 2B). All PAR-CLIP experiments reproducibly mapped Dicer binding (Figure 2A and 2B). PAR-CLIP reads and crosslink-induced T-to-C mutations cover the entire pre-miRNA stem-loop structure. AGO-bound sRNA reads align to both sides of the stem and form precise double stacks, representing the “mature” and “star” miRNA. Additional ~11 nt of base-paired RNA beyond the pre-miRNA reflect the requirements for Drosha processing (Han et al., 2006) of miR-17 and miR-40 (Figures 2A and 2B, inserts).

In contrast to sRNA sequencing, Dicer PAR-CLIP abundantly yielded reads spanning the loop region of miRNAs. This was further enhanced when the RNase T1 treatment was omitted (Figure S2A and S2B). Thus, our PAR-CLIP assay faithfully recovers Dicer binding to intact miRNA precursors.

To our surprise, Dicer binds not only pre-miRNAs. We mapped Dicer-binding sites to GENCODE, WormBase in *C. elegans*, and miRBase annotations. Only 321 binding sites map to 316 individual miRNAs in HEK293 cells and 100 binding

sites to 92 miRNAs in *C. elegans*. The majority of Dicer-binding sites map to a wide range of RNA transcripts, including not only other structural RNAs, such as tRNAs and snoRNAs, but also exonic and intronic mRNA regions (Figures 2C and 2D and Table S1). This is also reflected in the distribution of PAR-CLIP reads (Figures S2C and S2D).

Dicer-Binding Sites Explain the Majority of Expressed sRNA

We asked how many of the expressed sRNAs map to Dicer-binding sites. We sequenced AGO-loaded sRNAs in HEK293 cells by immunoprecipitation of stably expressed FLAG/HA-tagged AGO2 and AGO3. The two data sets were highly correlated (Figure S2E) and subsequently merged (referred to as AGOIP). AGOIP read counts were also highly correlated to normal 5'-monophosphate-specific (5'-P) sRNA sequencing, indicating that the majority of expressed sRNAs are indeed loaded into AGO (Figure S2F). Of the aggregate number of AGOIP reads from 18–26 nt length, 95% originate from human Dicer-binding sites, comparable to the 93% explained by miRBase (Figure 2E).

C. elegans expresses at least 27 Argonautes (Gu et al., 2009). We therefore relied on conventional sRNA sequencing. Here, 61% of 18–26 nt reads originate from nematode Dicer-binding sites, comparable to 60% explained by miRBase (Figure 2F). The previously published PAR-CLIP data sets of single-stranded RNA-binding proteins, human ELAVL1 (Lebedeva et al., 2011) and *C. elegans* GLD-1 (Jungkamp et al., 2011), were analyzed in the same manner as negative controls and explained less than 1% of the expressed sRNA. Worm-specific, triphosphorylated endo-siRNAs were sequenced independently and were found to be largely unassociated with Dicer-binding sites, consistent with previous results on their Dicer-independent production (Gu et al., 2009) (Figure S2G).

Dicer PAR-CLIP Identifies sRNA-Generating Loci with High Sensitivity

Because the pool of expressed sRNAs can be dominated by a few highly expressed miRNAs, we investigated the sensitivity of the Dicer PAR-CLIP as a function of sRNA expression. We considered all clusters of AGOIP reads that uniquely align to the genome as potential sites of sRNA production and selected those with a given minimal sRNA output. We then asked what fraction of these stacks overlap Dicer-binding sites, known miRNAs, or control sites (Figures 2G and 2H). Although Dicer-binding sites more often overlap with regions of low and intermediate sRNA expression than the miRBase annotation, the gap disappears for highly expressed sRNAs. Thus, although comparable fractions of total sRNA can be explained by either miRBase or Dicer binding, Dicer PAR-CLIP offers enhanced sensitivity for lowly expressed sRNAs.

We conclude that Dicer PAR-CLIP identifies loci of sRNA production with high sensitivity and that the majority of AGO-loaded or 5'-P sRNAs (at least 95% in HEK cells and at least 61% in *C. elegans*) are probably Dicer products. Furthermore, the control data sets for ELAVL1 and GLD-1 demonstrate that our analysis of HEK293 and worm PAR-CLIP data has a very low rate (<1%) of falsely reporting sites that generate sRNAs.

Dicer PAR-CLIP Discovers Hundreds of Additional miRNAs

Next, we searched for previously unidentified miRNAs in the Dicer PAR-CLIP data. We identified 1,978 human DICER1-binding sites with AGOIP read counts above background level (Figure S3A). Of these, 1,678 are not listed in miRBase (Figure S3A). We ran miRDeep2 (Friedländer et al., 2012) to predict additional miRNAs from the AGOIP data independently and subsequently intersected these with DICER1-binding sites. This analysis resulted in 212 (not present in miRBase) miRNA candidates supported by Dicer PAR-CLIP (Table S2). However, we also found many Dicer-bound loci with a clean “double-stack” profile of AGOIP reads, the signature of processed pre-miRNAs, which were only called by miRDeep2 after deactivating the hairpin scoring (Table S2). This scoring is meant to assess how likely the hairpin fold of a sequence occurred by chance, which reduces false positives but also sensitivity. We successfully validated Dicer-dependent processing for three out of three such miRNAs by an *in vitro* processing assay (Figure S3B), demonstrating that the PAR-CLIP evidence for *in vivo* interactions allows us to relax the miRDeep2 filtering. In total, we report 367 new miRNA candidates predominantly originating from introns (like known miRNAs), but also from a variety of other sources including 5' UTRs (Figures 3A and 3B).

In *C. elegans*, only two sites were identified as putative additional miRNAs (Figures S3C and S3D), which is not surprising considering the very deep sRNA profiling and smaller genome (Shi et al., 2013).

Additionally Identified miRNAs Are Dicer Dependent and Interact with mRNAs

To validate the miRNA candidates, not present in miRBase, we compared sRNA read counts from control cells to Dicer- or Drosha-depleted cells (Figures S3E–S3G). Our 5'-P sequencing data allowed us to quantify expression changes of 52 miRNA candidates, which were specifically and highly significantly downregulated upon Dicer knockdown (Figures 3C, S3H, and S3I).

Next, we screened publicly available AGO CLIP data for chimeric reads (Grosswendt et al., 2014) that would support direct miRNA:target interactions. For 20 of our miRNA candidates, chimeric reads with target 3' UTRs were found (Table S2). Although some indicate binding through seed complementarity, others show noncanonical miRNA:target interactions. For example, the miRNA derived from the 5' UTR of glutamate-ammonia ligase (GLUL) appears to bind to the 3' UTR of methyltransferase-like 8 (METTL8) by base-pairing within the central region of the miRNA (Figure 3E).

We further validated four out of four candidates by northern analysis. Dicer dependence is reflected either by reduction of mature miRNA or by accumulation of pre-miRNA in Dicer-depleted cells (Figure 3D). Finally, we employed a luciferase reporter with perfect matches to miRNA candidates. By this assay, three out of six candidates can regulate targets *in vivo* (Figure 3F). In summary, multiple independent assays support expression, Dicer dependency, and functionality of the additionally identified miRNAs.

Dicer Binds and Cleaves Structural RNAs

Known and additional miRNAs still account for only a fraction of the Dicer targets with sRNA output. We call these sites “active”

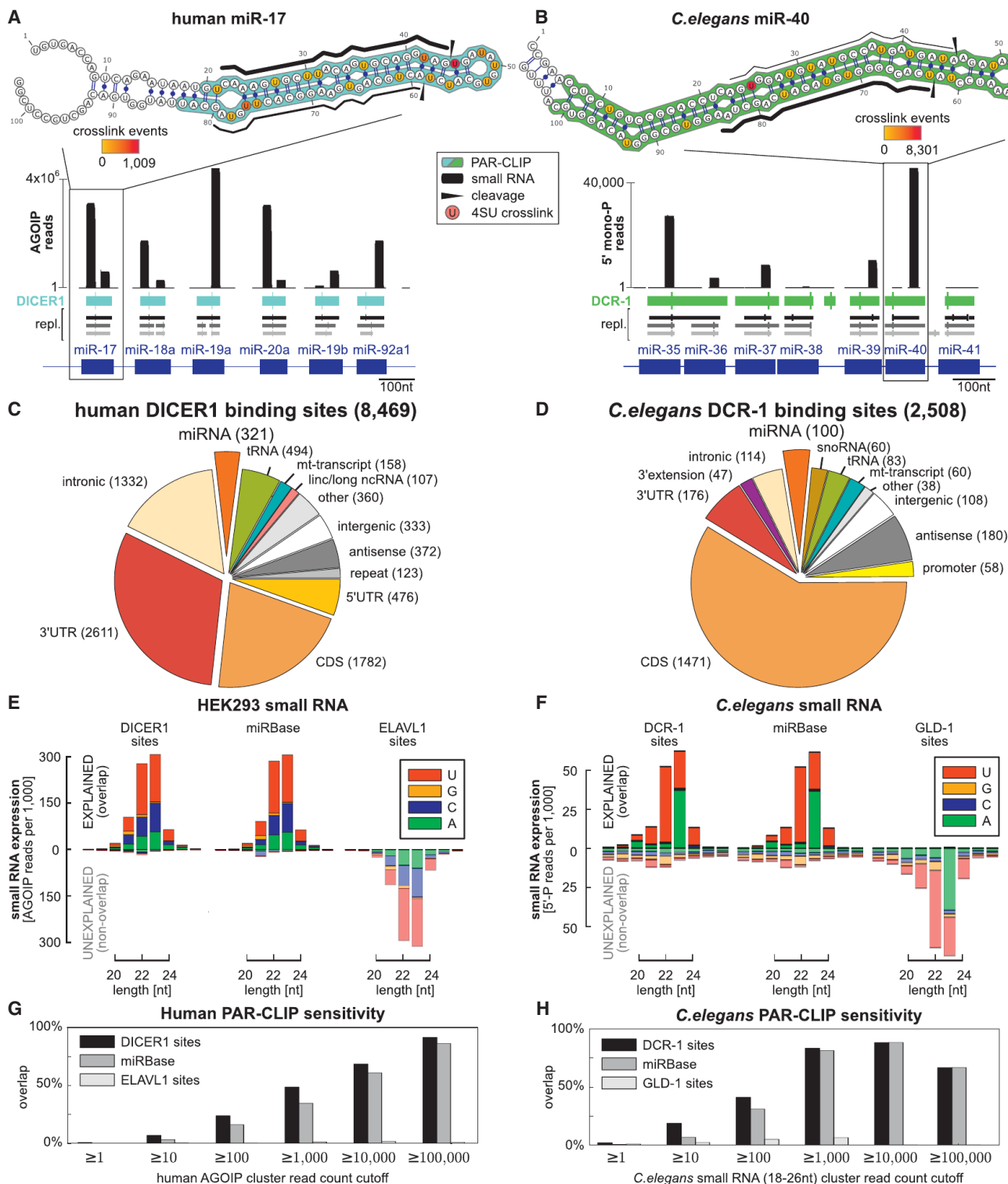


Figure 2. PAR-CLIP Recovers Known miRNAs and Explains the Majority of Expressed sRNA

(A and B) Dicer PAR-CLIP data for miRNA clusters: human miR-17–92 (A) and *C.elegans* miR-35–41 (B). PAR-CLIP-binding sites from three independent experiments are indicated as gray boxes. Consensus clusters in blue (human) or green (nematode). Human AGO2/3-IP sRNA reads and *C. elegans* 5'-P sRNA read coverage in black. The top insert shows details of the folded pre-miRNA secondary structure. PAR-CLIP read coverage in blue (human) and green (*C. elegans*). Crosslinks in yellow to red, proportional to induced nucleotide conversion frequency. sRNA coverage as black lines, with width indicating expression level.

(legend continued on next page)

(at least 10 AGOIP or sRNA reads) or “highly active” (human: >1,000 AGOIP reads, *C. elegans*: >100 sRNA reads) (Figures S4A and S4B). Figure 4A shows active Dicer binding to the H/ACA box snoRNA SNORA33, the 3' splice site of an alternative exon of Droscha mRNA, and the 5' UTR of glutamate-ammonia ligase (GLUL) (Figure S4C shows greater detail). All of these sites give rise to ample sRNAs, comparable to a medium-expressed miRNA. The *C. elegans* examples of atypical substrates include the C/D box snoRNA F30H5.4 and Y-RNA *ymn-1* (Figure S4D), where cleavage may generate different RNA fragments of ~50 nt length. Cleavage appears to frequently occur only on one side of the stem, suggesting that Dicer may primarily participate in its degradation. Expression of Dicer-bound structural RNAs indeed increased upon Dicer depletion, as shown by qPCR for tRNAs, snoRNAs, vault RNAs, or Y-RNAs in human (Figure S4E) and *C. elegans* (Figures S4F and S4G), independent of whether sRNA profiles match a clean double-stack profile that would resemble pre-miRNA-like processing.

Many Abundant sRNAs Are Processed with Low Precision

To study sRNA production for these substrates, we first computationally scanned binding sites for hairpin structures that could explain Dicer binding. These structures set the frame of reference to count sRNAs that align to both hairpin arms, to assess possible preferences for 5' or 3' arms, and to score the fidelity of sRNA processing by recording 5' variability of read stacks (Figure 4B and Experimental Procedures). The in silico hairpin folds are supported by PARS data, which captures secondary structures in vivo (Wan et al., 2014; Figure S4H).

The resulting picture reveals a large dynamic range (six orders of magnitude) of sRNA expression generated by diverse substrates (Figures 4C and S4I). The most abundant sRNAs originate from known miRNAs, whereas Dicer binding to mRNAs, on average (but with interesting exceptions like the 5' UTRs of GLUL or FLCN), gives rise to very little or no detectable sRNA. Expression of sRNAs from mRNAs did not correlate with mRNA expression (Figure S4J).

As reported before (Chiang et al., 2010), the more abundant sRNA preferentially arises from the 5' arm of the hairpin. This holds for known miRNAs and becomes more pronounced for the entire set of active sites (Figure 4D). Interestingly, expression level of sRNAs does not automatically correlate with processing fidelity (Figure 4E). Although known miRNA precursors are not only abundant, but also cut with the highest precision, the additional miRNAs and a subset of bound tRNAs and snoRNAs identified here are processed with comparable precision despite lower sRNA expression. On the other hand, abundant sRNAs from mitochondria are apparently produced with very low precision (Figures 4F and S4N).

There are examples of Dicer-bound structural RNAs, mostly tRNAs, which show clean stacks of AGO-loaded RNAs, but addi-

tional and more diverse sRNA profiles in 5' P sequencing (Figure 4F). Furthermore, upon Dicer knockdown, the sRNA output from most tRNAs is not substantially reduced (Figure S4K), whereas mitochondrial sRNAs even increase in abundance (Figures 4G and S4K). Taken together, Dicer binding to structural RNAs influences RNA stability, but concomitant sRNA production does not automatically indicate that dicing is also the source of the observed sRNA. Rather, abundant RNA species may be subject to multiple, different mechanisms of decay, including hydrolysis.

Mitochondrial Transcripts Are Regulated by Human and *C. elegans* Dicer

Because many Dicer-binding sites map to mitochondrial transcripts but display low fidelity of sRNA production, we examined the expression changes of mitochondrial transcripts upon Dicer depletion by qPCR. We reproducibly observed increased mitochondrial transcript levels after 2 days of DICER1 depletion in HEK293 cells (Figure 4H). The effect is even more prominent after 3 days (Figure S4L), which was confirmed by northern analysis (Figure S4M) and was similarly observed in DCR-1 RNAi worms (Figures S4O and S4P).

Dicer Is Associated with sRNAs Derived from Transcription Start Sites

As described before (Seila et al., 2008; Valen et al., 2011; Zamudio et al., 2014), some sRNAs originate from transcription start sites (TSS). We observe concomitant DICER1 binding around TSS of protein-coding genes (Figure S4Q), suggesting that promoter-associated sRNAs from both strands are, at least in part, DICER1 associated. In contrast, the AGOIP peak upstream of transcription termination and polyadenylation sites (PAS) does not coincide with Dicer binding (Figure S4R). This is consistent with recently published data describing PAS-associated sRNA as Dicer independent (Valen et al., 2011).

Reproducible Dicer Binding without Detectable sRNA Production

Although we were able to associate 86% of human Dicer-binding sites and 72% of *C. elegans* sites with local stem-loop folds, the majority of these Dicer-bound hairpins (73.4% and 84.5%) appeared to not give rise to detectable sRNA. These 5,349 human and 1,510 *C. elegans* “passive” sites are particularly enriched in coding sequences and 3' UTRs (Figures 5A and 5B).

Although we already demand that all binding sites are supported by at least two out of three independent PAR-CLIP replicates, we wanted to ensure that the passive sites indeed represent stable interactions with mRNAs. To this end, we performed independent Dicer immunoprecipitations without cross-linking or RNase treatment and assayed the bound RNAs by semiquantitative RT-PCR. In each case, the amplicons were set outside of the identified Dicer-binding sites, and the RT

(C and D) Annotation breakdown of binding sites in human (C) and *C. elegans* (D).

(E and F) Barplots of size and 5' nt distribution of sRNA reads. Top half, left to right: sRNA originating from Dicer-binding sites, miRBase miRNAs (pos. control), and ELAVL1/GLD1-binding sites (negative control). Bottom half: sRNA from remaining transcribed parts of the genome. (E) human, (F) *C. elegans*.

(G and H) Barplot showing the fraction of sRNA read stacks that overlap with Dicer-binding sites (black), miRBase miRNAs (dark gray), or ELAVL1/GLD1-binding sites as controls (bright gray), as a function of minimal sRNA read count in human (G) and *C. elegans* (H).

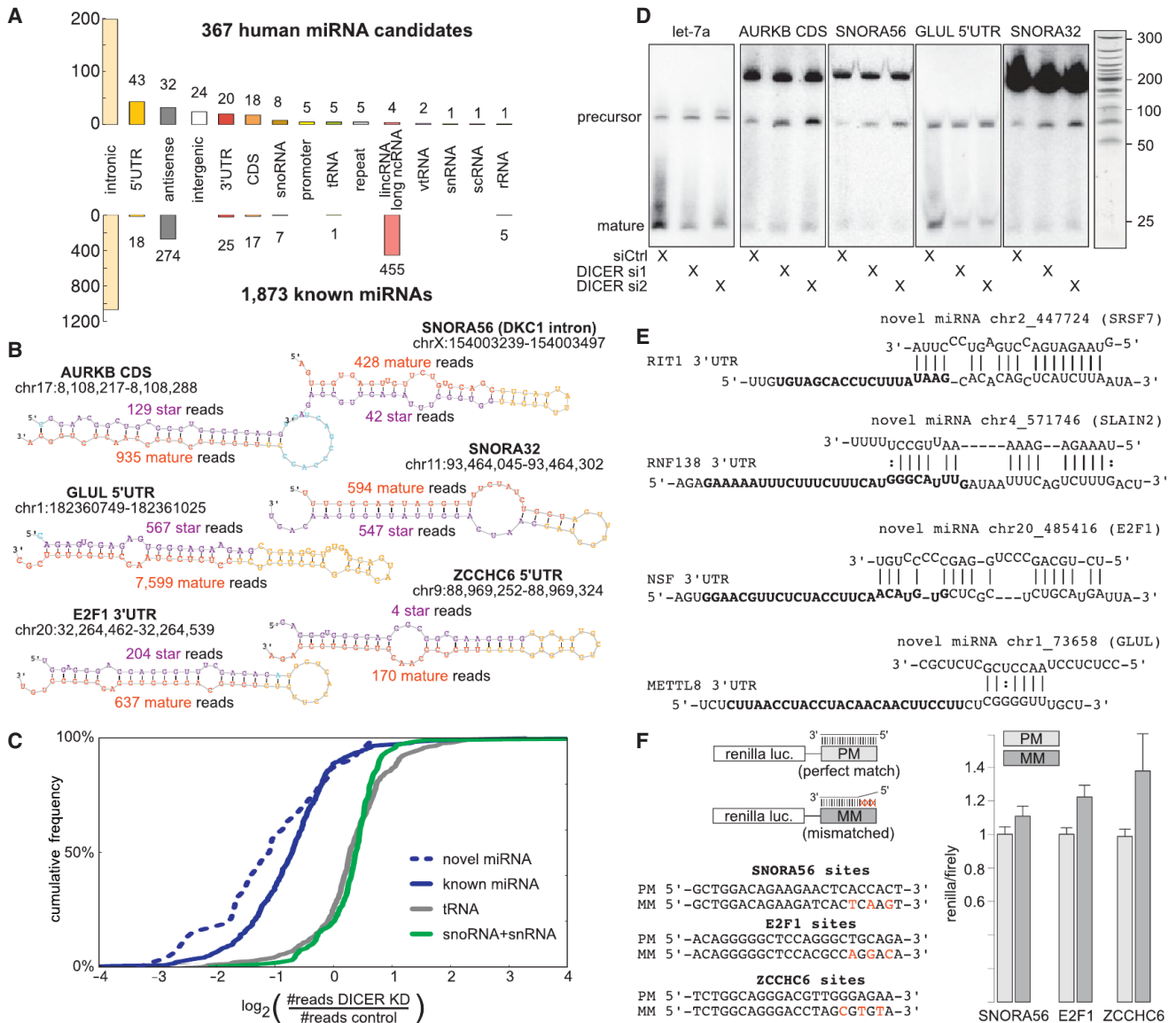


Figure 3. Identification and Validation of miRNA Targets

(A) Annotation breakdown of identified (top) and known miRBase (bottom) miRNAs.

(B) Examples of identified miRNAs of different human genomic origin. Reduced miRDeep2 plots show the precursor hairpin structure and the coverage of mature (red), star (violet), and loop (yellow) sequences.

(C) Cumulative distribution of changes in small RNA expression upon Dicer knockdown (siRNA 2). The distribution of log₂ fold changes of small RNA read counts of tRNAs (gray), snoRNAs, and snRNAs (green), known miRNAs (dark blue), and miRNAs identified in this analysis (blue, dashed). Known ($p < 1 \times 10^{-88}$, Mann-Whitney U, $n = 512$) and newly identified ($p < 1 \times 10^{-19}$, MWU, $n = 52$) are significantly more downregulated after Dicer depletion than tRNAs ($n = 513$).

(D) Northern blots of human let-7a (pos. control) and four miRNAs validate the ~22 nt mature product and either reduction of mature (let-7a, GLUL) or increase of precursor (AURKB, SNORA56, SNORA32) upon Dicer-depletion.

(E) Chimeric reads from AGO-CLIPs, containing miRNAs (top) and target fragments (below, fragments are extended; original chimeric part is bold) support in vivo interaction. Vertical bar: Watson-Crick pairing; colon: G-U wobble. Dashes indicate bulged nucleotides in the paired sequence.

(F) Luciferase reporter assay with perfectly matching (PM) or mismatched (MM) binding sites controlling Renilla expression. Barplot: Ratios of Renilla to firefly (control). Error bars indicate SD.

reaction was oligo-dT primed to specifically test for binding to intact mRNAs. By this assay, nine out of ten passive Dicer-binding events were validated in both human and worm (Figures 5C, 5D, S5A, and S5B), also with an antibody against endogenous DICER1 (Figure S5A).

In *C. elegans*, the most strongly bound passive target of DCR-1 in vivo is rncs-1 (Figure 5E). This lincRNA forms a long double strand that was previously described to bind DCR-1 with high affinity in vitro while being resistant to dicing due to secondary structures at the ends of the long stem (Hellwig and

Bass, 2008). The absence of abundant sRNAs from *mcs-1* argues for the absence of efficient dicing, consistent with the proposed role of *mcs-1* as a stable competitor for other Dicer substrates (Hellwig and Bass, 2008).

The transcript with the largest number of passive DICER1-binding sites in human is the mRNA encoding DICER1 itself. It is covered with 39 PAR-CLIP clusters, but like *mcs-1* in worm, these sites do not emit sRNAs (Figure 5F). Another strongly bound example is a binding site in the 3' UTR of vesicle-associated membrane protein-associated protein B/C (VAPB) (Figure 5G). The sites constitute imperfect stem-loop structures, similar to pre-miRNAs, but with larger bulges and lacking the distinct 11 nt stem segment that would be required for Drosha processing.

Passive Sites Can Interfere with Catalytic Dicer Activity

As *mcs-1* expression was shown to inhibit the catalytic activity of DCR-1 via sequestration of DCR-1 (Hellwig and Bass, 2008), we tested whether this is also possible with human passive sites.

We expressed four mRNA transcripts containing passive hairpins of different origin (DICER1, SLC2A1, TARBP2, VAPB) to high levels (in the range of GAPDH) in HEK293 cells (Figure S5C). Expression of four out of four passive hairpins reproducibly reduced the levels of endogenously expressed miRNAs by 20%–30% in comparison to constructs bearing mutations that disrupt the secondary structure (Figure 5H), suggesting that these interactions are sufficiently strong to functionally sequester Dicer protein in vivo. Dicer protein levels were not affected (Figure S5C). Taken together, our data show that Dicer stably and reproducibly binds to specific stem-loop structures inside of intact mRNAs without dicing.

Passive Dicer Targets Are Functionally Linked to RNA Granules

Mammalian and *C. elegans* Dicer is essential for the germline (Knight and Bass, 2001; Murchison et al., 2007). A hallmark of animal germ cells are RNA granules, perinuclear aggregates of RNA and protein. In *C. elegans*, Dicer is present in and required for assembly of these granules (Figure S5D and Beshore et al., 2011). As Dicer is a large protein with many protein interactors and passive binding to mRNAs does not involve RNase activity, we hypothesized that Dicer may additionally play a role in aggregating bound mRNAs to RNA granules. We intersected lists of genes known to be required for the proper formation of RNA granules with Dicer targets. To qualify as a passive Dicer target, we conservatively demand that a gene must not contain a single Dicer-binding site with sRNA output exceeding background levels (Experimental Procedures). We find that both human ($p < 0.015$ Fisher's exact test) and worm ($p < 0.003$ for P bodies, $p < 2.3 \times 10^{-9}$ for P granules) passive target transcripts of Dicer are statistically significantly enriched for genes associated with granules (Updike et al., 2011) (Figures 5I and 5J). For many of these genes, it has been demonstrated that both, their mRNAs and proteins, are localized to granules (Schisa et al., 2001). Of note, the enrichment of P-body-associated genes among Dicer targets in HEK293 cells is almost entirely due to passive targets and becomes even more significant for the most strongly bound passive targets with most crosslink events. In contrast, active

targets are not enriched for granules. The association of passive Dicer binding with RNA granules suggests a role in mRNA localization. Indeed, our FLAG-tagged DCR-1 protein recapitulates the granular localization of wild-type DCR-1 in the *C. elegans* germline (Figures 5K and S5E and Beshore et al., 2011).

Human passive targets of Dicer are also significantly ($p < 2.8 \times 10^{-5}$) enriched for 3' UTR targets of Staufen (Ricci et al., 2014), presumably due to the presence of double-stranded structures in their common targets. However, whereas Staufen binds to long stems (Ricci et al., 2014), Dicer binding in HEK293 cells seems restricted to the tip of the structure and requires contact to the loop region of hairpins (Figure S6B).

Homologous, Passive Targets in Human and *C. elegans*

We find 82 genes with homology between human and worm to be passive Dicer targets in both species. An interesting example is the 3' UTR of the germline helicase CGH-1 (Figure 5L): the human ortholog DDX6 is also bound by DICER1 via its 3' UTR (Figure 5M). Although the comparison of an animal and a human cell line across ~550 Mya of evolution does not warrant to judge conservation of function, we find the prominent appearance of ancient genes linked to RNA granules and the germline intriguing and report passive, homologous targets in Table S3.

Active and Passive Sites Differ in Secondary Structure

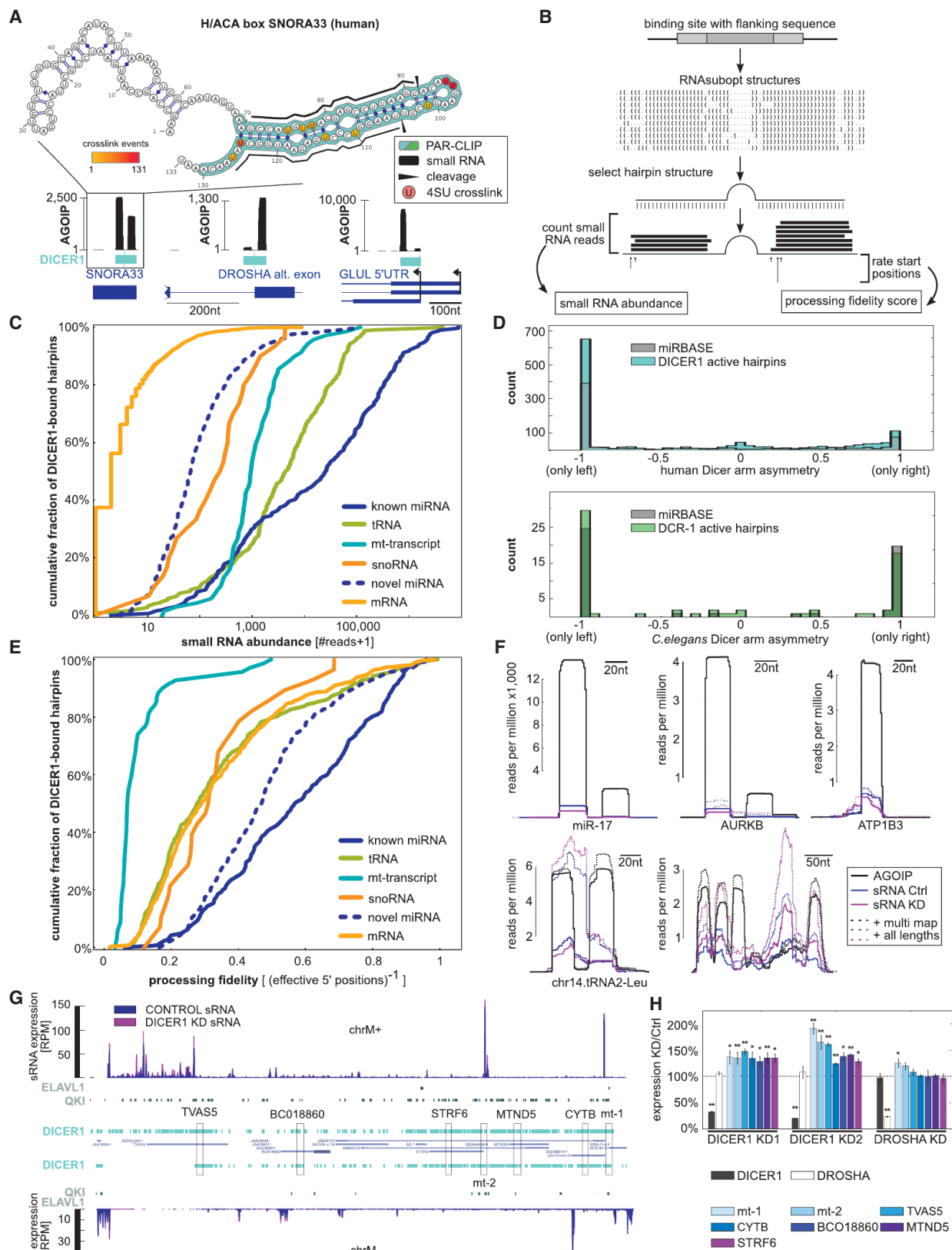
To characterize the differences between active and passive sites, we grouped Dicer sites by their activity (sRNA output level). Taking the center of the loop as a point of reference, we computed average profiles of structure, crosslinking, and other features as a function of position within the stem loop.

Known miRNAs have a characteristic fold with an unpaired loop region, flanked by paired bases within the stem of the pre-miRNA. Importantly, separated by a bulge, the stem extends for another ~11 nt beyond the pre-miRNA boundaries, which allows processing by DROSHA/DGCR8 (Han et al., 2006). This profile is observed for active sites (Figures 6A and 6B), including the miRNA candidates identified in our analysis, and is less pronounced for sites with lower activity. Among the active sites, base-pairing stability correlates with sRNA output, which, however, appears to be optimal for intermediate folding energies (Figure S6C). In contrast, base-pairing within passive sites decays steadily to background levels, beyond ~15 bp of stem. Figures S6D–S6G show additional profiles of G/C content and bulges.

Passive Sites Preferentially Crosslink within the Loop, Active Sites at the Ends of Precursors

If passive sites are indeed within intact mRNA transcripts, in contrast to pre-miRNA-like hairpins, no 5' and 3' ends should be accessible to Dicer. This difference could be reflected by the frequency of crosslink-induced transitions along the stem-loop structure because crosslinking requires close spatial proximity (Hafner et al., 2010).

Crosslinks between Dicer and HEK293-expressed miRNAs peak in three distinct regions: within the loop and at the 5' and 3' ends of the pre-miRNA (Figure 6C). The loop can interact with the Dicer helicase domain (Taylor et al., 2013), whereas



(legend on next page)

the ends are coordinated by the 5' pocket and PAZ domain, which is required for "setting the ruler" and efficient dicing (MacRae et al., 2007; Park et al., 2011). This pattern is observed for all active sites, supporting that active sites are bound as pre-miRNA-like structures cleaved out of larger transcripts. In contrast, passive sites lack the peaks at precursor-size distance (Figure 6C). In *C. elegans*, the same difference between active and passive sites at the precursor termini is observed. However, active sites display much less crosslinking within the loop region (Figure 6D), consistent with the dispensability of the helicase domain for miRNA production in *C. elegans* (Welker et al., 2011).

Passive Sites Are Not Engaged by DGCR8 and Are Not Cleaved

For humans, we additionally interrogated DGCR8 HITS-CLIP data (Macias et al., 2012) and 5'-monophosphate sequencing of endonucleolytic RNA cleavage products (Karginov et al., 2010). Active sites and known miRNAs display prominent shoulders of DGCR8 HITS-CLIP signal, flanking the pre-miRNA and indicating processing by the Drosha/DGCR8 complex. This signal is absent from passive sites (Figure 6E). Moreover, binding of the Dicer-interacting proteins PACT and TARBP2 (Goodarzi et al., 2014) was only detected for loci with sRNA output (Figures S6H–S6J). Consistently, only active sites, but not passive sites, are supported by cleavage products aligning to either the pre-miRNA itself or the 3' fragments of the primary transcript.

We conclude that the passive Dicer-binding sites observed in human and nematode represent local stem-loop structures, which structurally differ from typical miRNAs and are consequently not cleaved by Drosha or Dicer. Rather, they constitute structural elements within intact transcripts.

Functional Consequences of Dicer Binding and Dicing

To assess the significance and biological consequences of widespread Dicer binding to mRNAs, we performed Dicer knockdowns in HEK293 cells and *C. elegans*, followed by mRNA sequencing. To limit indirect effects due to reduced miRNA production, we chose the earliest time point at which Dicer protein levels were significantly reduced while miRNA levels were stable. As an additional control, we knocked down Drosha (Figures S7A and S7C–S7F).

In *C. elegans*, miRNA expression changes during development, and Dicer depletion causes pleiotropic effects, including sterility. Our best attempt to study the impact of Dicer binding on mRNAs therefore consists of choosing a developmental stage when Dicer can be significantly depleted (Figures S7B, S7G, and S7H) while the animals are viable. We used the *fem-1 (hc17)* strain (inducible sterility) to obtain sterile control animals.

As the sequencing-derived log fold changes correlated well with qPCR measurements on independent biological replicates (Figure S7M), we compared the distribution of changes between different groups of transcripts. Overall, the observed changes at early time points are small. However, in both HEK293 cells and *C. elegans*, we observed statistically significant effects on Dicer target transcripts. mRNAs containing active Dicer-binding sites were significantly stabilized upon Dicer depletion, whereas passive targets overall were more destabilized in both worms and cells (see Figure S7N for mRNA decay). The most strongly bound passive targets, as well as the homologous passive targets of DICER1 and DCR-1, show the strongest effects (Figures 7A, S7I, and 7B).

In contrast, Drosha depletion had almost no effect on passive targets in HEK293 cells (Figures S7J and S7K), ruling out that the observed regulation stems from global loss of miRNAs. In *C. elegans*, where results strongly depended on the developmental stage, Drosha depletion stabilized both passive and active targets (Figure S7L).

DISCUSSION

The rapid increase of sequencing data has led to the discovery of many sRNAs that have been linked to Dicer activity. However, sRNA data capture only the endpoint of a cascade of processing events on a background of degradation products, and the indirect identification of Dicer binding from sRNA requires assumptions about miRNA-like processing. Here, we biochemically identified thousands of in vivo Dicer-binding sites in a human cell line and *C. elegans*.

Although we recovered known miRNAs with high sensitivity, these account for only a fraction of our data. We identified and validated many additional miRNAs. Most of these are lowly expressed and would not have been detected without direct evidence for Dicer binding.

Figure 4. Dicer-Binding Sites in Different RNA Substrates Display a Wide Dynamic Range of sRNA Expression and Processing Fidelity

(A–F) Dicer binding to noncanonical structural RNA: snoRNA SNORA33, insert shows structure as in Figure 2A, 3' splice site of DROSHA mRNA, 5' UTR of GLUL mRNA in HEK293 cells. Dark blue: exon/intron structure of bound transcript; light blue: Dicer-binding site, AGOIP coverage in black.

(B) Schematic of the local hairpin search. Binding sites (dark gray) are padded with flanking sequence (light gray) and are folded with RNAsubopt. The best hairpin is selected and intersected with aligned AGOIP/sRNA reads. Numbers and positions of alignments used to score output, left/right arm asymmetry, and fidelity of sRNA processing are shown.

(C) Cumulative frequency of AGOIP read counts for different categories of Dicer-binding sites in HEK293 cells.

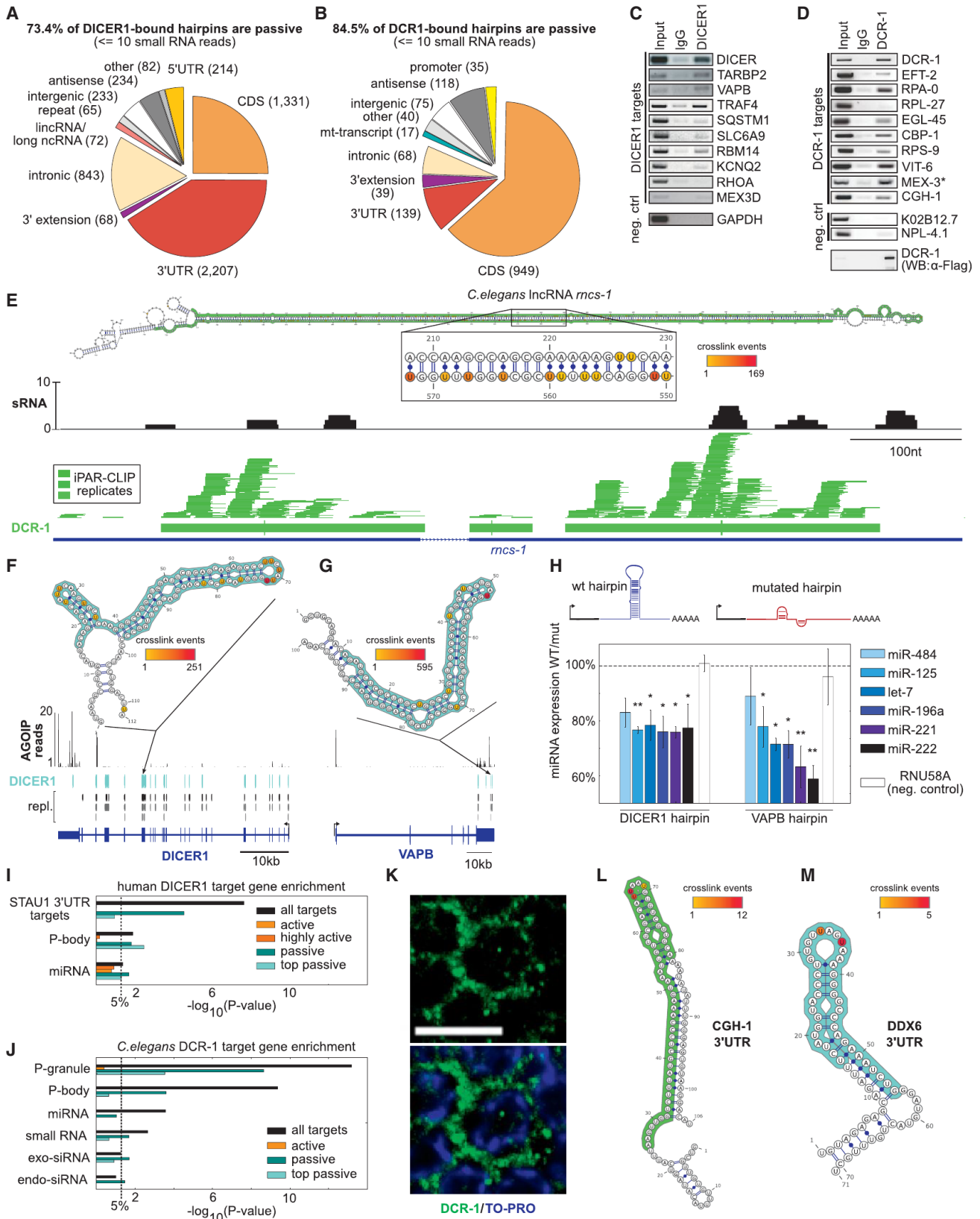
(D) Small RNA production is biased toward the left arm. Histogram of sRNA read coverage asymmetry across all sRNA precursors in human (top) and *C. elegans* (bottom). Dicer-bound active hairpins (bright blue and green, respectively) and miRBase annotated human and *C. elegans* pre-miRNAs (dark gray).

(E) Cumulative frequency of AGOIP read stack fidelity (0 for randomly scattered reads, 1 for a single 5' start position for all reads) for different categories of Dicer-binding sites in HEK293 cells.

(F) Examples of AGO-loaded small RNA sources. Black: AGOIP reads, blue: 5'P sRNA control, purple: 5'P sRNA DICER1 KD2. Uniquely aligning reads 18–25 nt. Dashed lines: all reads.

(G) Dicer (blue boxes) and neg. control QKI and ELAVL1 (gray boxes) binding along mitochondrial transcripts (middle, blue) from the plus strand (top) or minus strand (bottom). sRNA output in blue (control) and purple (DICER1 KD with siRNA2). Amplicons for qPCR are marked by rectangles.

(H) Barplot with qPCR changes of mitochondrial transcripts upon Dicer KD. Error bars represent SEM. * $p < 5\%$, ** $p < 1\%$, two-sided, two-sample t test on delta Ct, unequal variance, $n = 3$.



(legend on next page)

Furthermore, the intersection of Dicer binding, sRNA sequencing, and AGOIP allowed us to investigate Dicer interactions with structural RNAs. We found not only unusual patterns of dicing, but also Dicer-independent, AGO-loaded sRNAs, presumably generated by other modes of processing or degradation. As Dicer-bound structural RNAs accumulated upon Dicer depletion (Figures S3F and S3G), regardless of the fidelity of their processing, this opens the possibility that, rather than producing *trans*-acting miRNAs, Dicer may contribute to or interfere with the degradation of many structural RNAs. Dicer is required for germline maintenance and early embryogenesis (Knight and Bass, 2001; Murchison et al., 2007), and its expression changes during differentiation. This observation could, in part, explain why the pool of tRNAs differs between proliferating and differentiated cells (Gingold et al., 2014).

To our surprise, depletion of Dicer, on average, increased the levels of mature mRNAs hosting active sites, suggesting that there can be some form of feedback to DROSHA/DGCR8 processing. Components of the sRNA pathway are known to auto- and cross-regulate each other (Bennasser et al., 2011; Han et al., 2009), and many, including three AGOs, TARBP2, and DROSHA, are also targeted by DICER1. Indeed, Dicer and Drosha levels were not independent in perturbation experiments. However, at the early time point in Figures 7A and 7B, Drosha levels were unaffected.

Based on studies in nematodes, it has been suggested that long dsRNA represents the archetypal substrate for Dicer, linked to its supposed ancient role in defense against dsRNA viruses or genome-invading elements (Sarkies and Miska, 2013). However, it appears that binding of human, somatic Dicer requires contact to the loop region, presumably utilizing the helicase domain, as we observe prominent crosslinking to hairpin loops. Indeed, we find almost no long dsRNA binding in HEK293 cells (Figures S4S, S6A, and S6B), in line with other findings that somatic mammalian cells lack endo-siRNA biogenesis (Nejepinska et al., 2012).

Our unbiased look at Dicer-binding sites pictures a continuum of sRNA expression from diverse substrates and surprisingly encompasses a large set of non-diced, passive binding sites that were previously undetectable. This class of passive binding sites, predominantly residing in mRNAs, can be stably bound

without endonuclease activity. This observation agrees with previous reports indicating that Dicer is not able to efficiently process RNA if no free ends are available (Fukunaga et al., 2014; Zhang et al., 2002) and explains the resistance of the lncRNA *mcs-1* against dicing (Hellwig and Bass, 2008).

Our data argue that passive sites are not transient interactions and that they can stabilize expression of their host transcripts. We note that these effects were strongest for mRNAs that were most strongly crosslinked or are also passive targets in the other species.

What could be the function of passive Dicer sites besides generally stabilizing target RNAs? First, we have shown that high expression of passive sites can sequester Dicer protein from other target transcripts and can interfere with its catalytic activity, analogous to lncRNA *mcs-1* in *C. elegans*. In these experiments, we expressed a few thousand passive sites per cell. It is not trivial to estimate the total copy number of all passive sites, but according to our data, a high number of passive sites may be naturally present. These sites could have a buffering function and may compete with active sites for Dicer binding. Second, the proteins encoded by passively bound mRNAs can immediately interact with Dicer, as they are translated. This would be a mechanism for efficient buildup of protein complexes. In support of this idea, several mRNAs with passive sites both in human and worm encode known Dicer interactors (DRH-1, LIN-41, TARBP2, DHX9). Third, many passively bound mRNAs encode RNA-granule-associated genes, and we show that Dicer protein itself is granularly localized in the *C. elegans* germline. Taken together with the observation that Dicer is required for RNP granule formation (Figure S5D and Beshore et al., 2011), this indicates that passive binding may be important for RNA localization and assembly of RNA-protein complexes. This finding could explain the diverse although overall weak effects observed upon DICER1 knockdown in HEK293 cells and suggests that germ cells, forming prominent RNP structures, are a more suitable system to investigate this aspect of Dicer biology. Along the same line, the *fem1* strain of *C. elegans* forms large and prominent P granules in its arrested oocytes (Schisa et al., 2001), and the observed changes of passive targets upon DCR-1 depletion were more pronounced in these animals than in the wild-type (data not shown). Independently supporting

Figure 5. Most mRNA Sites Are Not Substrates for Dicer Cleavage

(A and B) Pie charts of genomic annotation of passive binding sites for human (A) and *C. elegans* (B).

(C and D) Validation of catalytically passive Dicer targets by RIP RT-PCR. RT-PCR on RNA from Dicer IP validated nine out of ten tested in HEK293 cells (C) and nine out of ten in *C. elegans* (D). RT-PCR for abundant GAPDH mRNA served as negative controls for human. Mouse IgG IP and RT-PCR for three abundant transcripts: K02B12.7, NPL4.1 as negative control in *C. elegans*. RT-PCR reaction primed with oligo-dT primer and amplicon set outside of PAR-CLIP-binding sites. *MEX-3 has low PAR-CLIP coverage, is excluded from the strict consensus list, and was included here for comparison.

(E) *C. elegans* highly double-stranded lncRNA *mcs-1* is densely bound, but not efficiently cleaved by DCR-1. Top: secondary structure as in Figures 1D and 1E with zoomed region to show crosslinks. Bottom: sense (antisense) aligned distinct PAR-CLIP reads are indicated as green lines above (below) the gene structure. Shades of green for replicates. Thick blue lines: exons, green boxes: consensus binding sites. sRNA read coverage of *mcs-1* locus in black.

(F and G) Examples of passive Dicer-binding sites in HEK293 cells: DICER1 mRNA (F), 3' UTR of VAPB mRNA (G). Dicer binding sites: blue boxes. AGOIP sRNA reads in black. Top inserts show predicted folding of Dicer-bound hairpins overlaid with PAR-CLIP data as in Figures 2A and 2B.

(H) Sponge assay. Barplots with the ratio of miRNA levels in cells expressing passive hairpins relative to mutated hairpins. Error bars represent SEM. * $p < 5\%$, ** $p < 1\%$, two-sided, two-sample t test, unequal variance, $n = 3$.

(I and J) Barplots of $-\log_{10} p$ value (one-sided Fisher's exact test) for target enrichment of human (I) and *C. elegans* (J) Dicer. Vertical line demarcates $p = 5\%$.

(K) FLAG-tagged DCR-1 expression in *C. elegans* germline RNP granules. Immunohistochemistry was performed on extruded gonads from adult worms using monoclonal anti-FLAG (green) antibody. Overview and higher-magnification view of adult germline are provided (Figure S5E). Cell nuclei were counterstained with TO-PRO (blue). Scale bar, 10 μm .

(L and M) Passive binding sites in the homologous targets conserved germline helicase-1 (CGH-1, *C. elegans*) and DEAD box RNA-helicase 6 (DDX6, human).

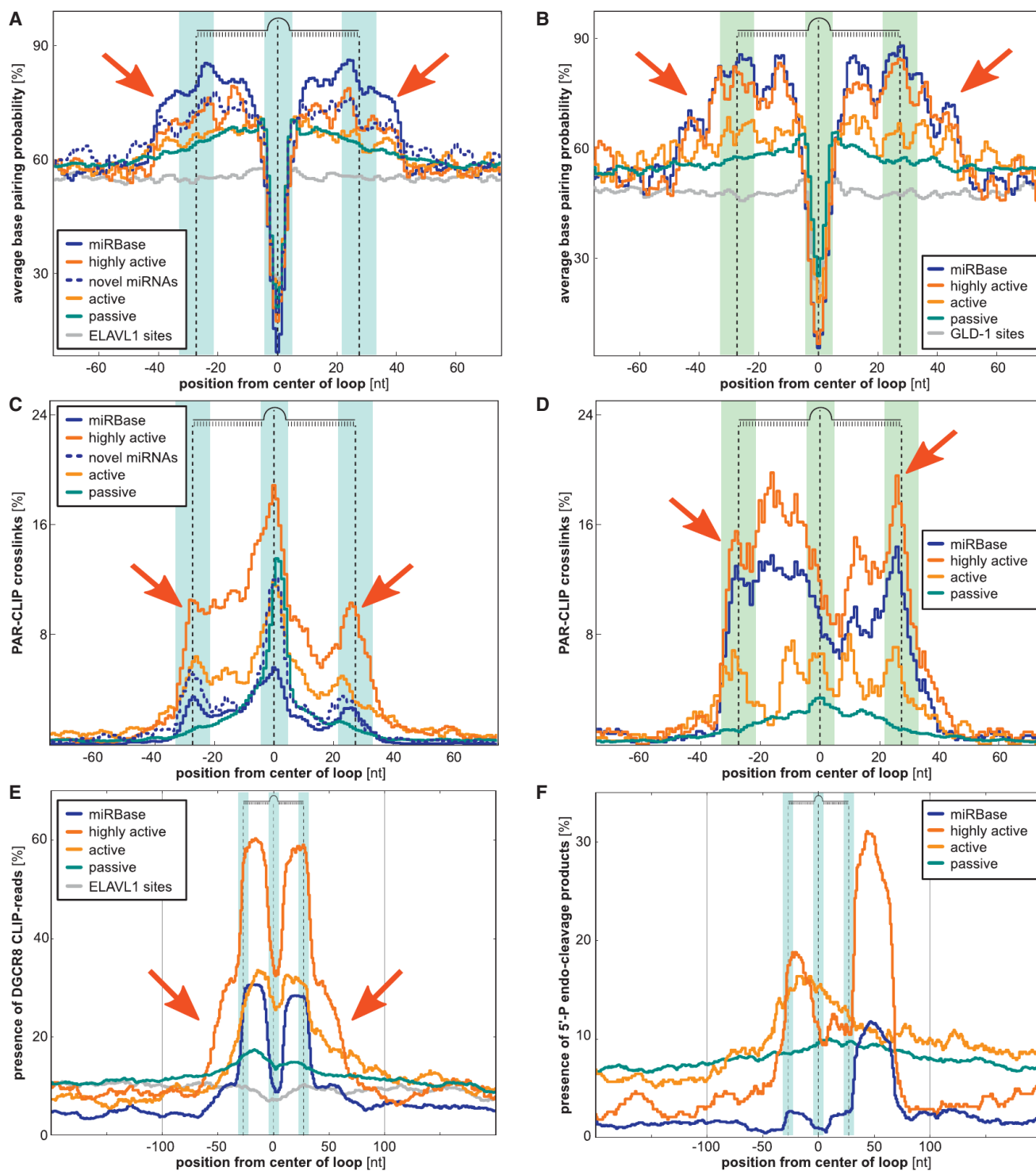


Figure 6. Secondary Structure and Dicer Crosslinking as a Function of sRNA Expression

(A and B) Average base-pairing probability at each nucleotide position in fixed windows around the loop center for human DICER1 (A) and *C. elegans* DCR-1-binding sites (B). Catalytic activity levels correspond to sRNA read count: highly active >1,000 (>100), active > 10, passive < 10 for human (*C. elegans*). ELAVL1/GLD-1 (single-stranded) binding sites as negative, miRBase miRNAs as positive control. Arrows point at the additional 11 nt of stem, characteristic for DROSHA/DGCR8 substrates.

(C and D) Average occurrence of crosslink mutations at each nucleotide position in fixed windows around the loop for human DICER1 (C) and *C. elegans* DCR-1 bound hairpins. Activity levels as above. Arrows point at the peaks of crosslinking at the 5' and 3' ends of precursor structures, absent from passive sites.

(legend continued on next page)

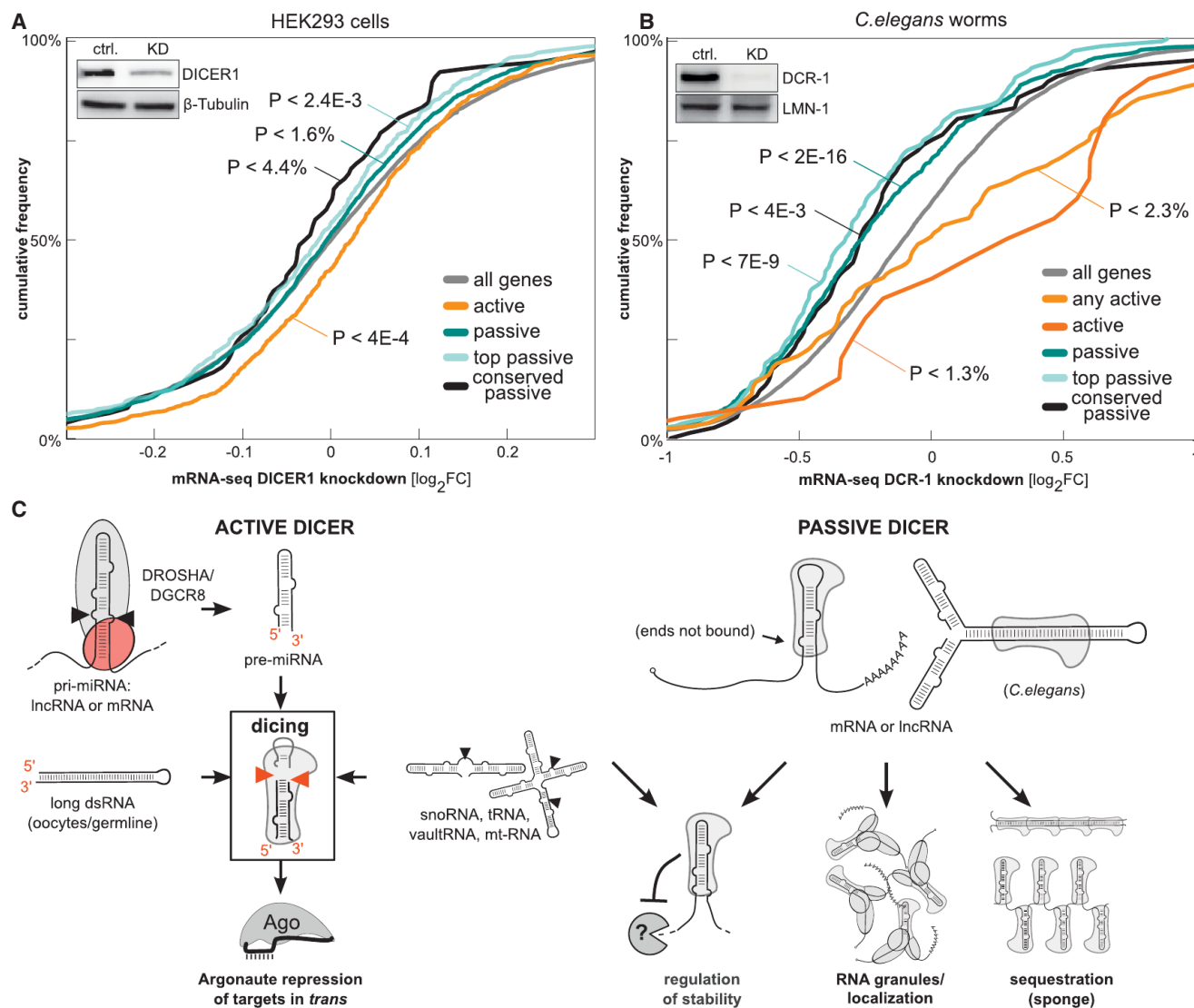


Figure 7. Functional Consequences of Dicer Binding and Dicing

(A and B) Cumulative frequencies of mRNA \log_2 fold changes for all detected genes (gray), catalytically active (orange/red), and passive targets (blue, most strongly bound: bright blue), as well as a conserved set of Dicer targets (black) in HEK293 cells (A) and *C. elegans* (B). Inserts: western blot validation of Dicer knockdowns. HEK293: active $p = 3.95 \times 10^{-4}$ ($n = 753$), passive $p = 1.586 \times 10^{-2}$ ($n = 2326$), top passive $p = 2.349 \times 10^{-3}$ ($n = 456$), conserved passive $p = 4.337\%$ ($n = 78$), all genes: $n = 10,007$. *C. elegans*: any active $p = 2.287 \times 10^{-2}$ ($n = 107$), active $p = 1.233\%$ ($n = 20$), passive $p = 1.695 \times 10^{-16}$ ($n = 804$), top passive $p = 7.412 \times 10^{-9}$ ($n = 165$), conserved passive $p = 4.269 \times 10^{-3}$ ($n = 75$), all genes: $n = 8501$. All p values by double-sided Mann-Whitney U.

(C) Model of Dicer-RNA interactions and function.

the idea that passive Dicer binding may influence mRNA localization, Dicer was recently found to be a shuttling protein (Doyle et al., 2013), and both human and *C. elegans* Dicer interact with nuclear pore complexes (Ando et al., 2011).

In summary, we mapped transcriptome-wide Dicer binding in human and *C. elegans*, identified and validated numerous

additional miRNAs, disentangled the contribution of Dicer to the degradation of many structural RNAs, and unveiled passive binding of Dicer with possible functions outside of sRNA pathways (Figure 7C). Also, the microprocessor components DGCR8/Pasha and DROSHA have recently been shown to have miRNA-independent cellular functions (Gromak et al.,

(E) Average presence of at least one DGCR8 HITS-CLIP (Macias et al., 2012) read for each nucleotide position in fixed windows around the loop for human DICER1-bound hairpins. Activity levels as above. Arrows point at the additional shoulders at the base, outside of pre-miRNAs, absent from passive sites.

(F) Average presence of at least one endocleaved 5'-P bearing sequencing read (Karginov et al., 2010) for each nucleotide position in fixed windows around the loop for human DICER1-bound hairpins. Activity levels as above. Drosha/DGCR8 processing produces one peak starting at the 5' end of the pre-miRNA and one at the remaining 3' fragment. Both are absent from passive sites.

2013; Macias et al., 2012). Altogether, this indicates that the versatile functions of these ancient RBPs may have been overshadowed by a focus on miRNAs and may need to be carefully re-examined.

EXPERIMENTAL PROCEDURES

Standard molecular biology techniques and lists of reagents as well as computational analyses are described in the Extended Experimental Procedures.

PAR-CLIP and iPAR-CLIP HEK293 Cells

PAR-CLIP on HEK293 cells stably expressing FLAG/HA-DICER1 was performed as described (Hafner et al., 2010). In vivo PAR-CLIP was performed on DCR-1::FLAG rescue strain (BB92; dcr-1(ok247)III,uuEx18) as described previously (Jungkamp et al., 2011).

AGO-Associated sRNA Cloning and Sequencing

Immunoprecipitation of FLAG/HA-tagged AGO2/3 cells was performed with FLAG magnetic beads. RNA was isolated, ligated to 3' and 5' adapters, reverse transcribed, and PCR amplified and sequenced.

sRNA Sequencing

sRNA sequencing was performed from 5 µg total RNA according to the standard Illumina sRNA library preparation protocol. For 5' triphosphate sRNA sequencing, RNA was treated with 5' polyphosphatase before library preparation.

ACCESSION NUMBERS

Sequencing data have been deposited in the GEO database under GSE55333.

SUPPLEMENTAL INFORMATION

Supplemental Information includes Extended Experimental Procedures, seven figures, and four tables and can be found with this article online at <http://dx.doi.org/10.1016/j.cell.2014.10.040>.

AUTHOR CONTRIBUTIONS

A.R.-W., M.J., and Y.M. contributed equally to this work. A.R.-W. and Y.M. designed and performed experiments, supervised by M.L. and N.R. A.R.-W. performed the *C. elegans* DCR-1 PAR-CLIP and the majority of validation experiments, assisted by M.H. Y.M. performed all of the human PAR-CLIPs, AGO-IPs, and validation experiments. M.J. carried out all of the computational analyses, supervised by N.R. N.R., M.L., M.J., A.R.-W., and Y.M. wrote the paper.

ACKNOWLEDGMENTS

The authors would like to thank all members of the Rajewsky lab for discussion and support. We thank T. Duchaine and W. Filipowicz for discussion and advice on the manuscript. We would like to thank H. Tabara for the anti DCR-1 antibodies. M.J. received a fellowship of the DFG-funded GRK1772. The group of M.L. was supported by the BMBF and Senate of Berlin. Y.M. was funded by DAAD fellowship.

Received: February 24, 2014

Revised: August 18, 2014

Accepted: October 3, 2014

Published: November 20, 2014

REFERENCES

Ando, Y., Tomaru, Y., Morinaga, A., Burroughs, A.M., Kawaji, H., Kubosaki, A., Kimura, R., Tagata, M., Ino, Y., Hirano, H., et al. (2011). Nuclear pore complex protein mediated nuclear localization of dicer protein in human cells. *PLoS ONE* 6, e23385.

Bartel, D.P. (2009). MicroRNAs: target recognition and regulatory functions. *Cell* 136, 215–233.

Bennasser, Y., Chable-Bessia, C., Triboulet, R., Gibbings, D., Gwizdek, C., Dargemont, C., Kremer, E.J., Voinnet, O., and Benkirane, M. (2011). Competition for XPO5 binding between Dicer mRNA, pre-miRNA and viral RNA regulates human Dicer levels. *Nat. Struct. Mol. Biol.* 18, 323–327.

Beshore, E.L., McEwen, T.J., Jud, M.C., Marshall, J.K., Schisa, J.A., and Bennett, K.L. (2011). *C. elegans* Dicer interacts with the P-granule component GLH-1 and both regulate germline RNPs. *Dev. Biol.* 350, 370–381.

Castellano, L., and Stebbing, J. (2013). Deep sequencing of small RNAs identifies canonical and non-canonical miRNA and endogenous siRNAs in mammalian somatic tissues. *Nucleic Acids Res.* 41, 3339–3351.

Chiang, H.R., Schoenfeld, L.W., Ruby, J.G., Auyeung, V.C., Spies, N., Baek, D., Johnston, W.K., Russ, C., Luo, S., Babiarz, J.E., et al. (2010). Mammalian microRNAs: experimental evaluation of novel and previously annotated genes. *Genes Dev.* 24, 992–1009.

Doyle, M., Badertscher, L., Jaskiewicz, L., Güttinger, S., Jurado, S., Huginschmidt, T., Kutay, U., and Filipowicz, W. (2013). The double-stranded RNA binding domain of human Dicer functions as a nuclear localization signal. *RNA* 19, 1238–1252.

Ender, C., Krek, A., Friedländer, M.R., Beitzinger, M., Weinmann, L., Chen, W., Pfeffer, S., Rajewsky, N., and Meister, G. (2008). A human snoRNA with microRNA-like functions. *Mol. Cell* 32, 519–528.

Feng, Y., Zhang, X., Graves, P., and Zeng, Y. (2012). A comprehensive analysis of precursor microRNA cleavage by human Dicer. *RNA* 18, 2083–2092.

Friedländer, M.R., Mackowiak, S.D., Li, N., Chen, W., and Rajewsky, N. (2012). miRDeep2 accurately identifies known and hundreds of novel microRNA genes in seven animal clades. *Nucleic Acids Res.* 40, 37–52.

Friedländer, M.R., Lizano, E., Houben, A.J., Bezdan, D., Báñez-Coronel, M., Kudla, G., Mateu-Huertas, E., Kagerbauer, B., González, J., Chen, K.C., et al. (2014). Evidence for the biogenesis of more than 1,000 novel human microRNAs. *Genome Biol.* 15, R57.

Fukunaga, R., Han, B.W., Hung, J.H., Xu, J., Weng, Z., and Zamore, P.D. (2012). Dicer partner proteins tune the length of mature miRNAs in flies and mammals. *Cell* 151, 533–546.

Fukunaga, R., Colpan, C., Han, B.W., and Zamore, P.D. (2014). Inorganic phosphate blocks binding of pre-miRNA to Dicer-2 via its PAZ domain. *EMBO J.* 33, 371–384.

Gingold, H., Tehler, D., Christoffersen, N.R., Nielsen, M.M., Asmar, F., Kooistra, S.M., Christophersen, N.S., Christensen, L.L., Borre, M., Sørensen, K.D., et al. (2014). A dual program for translation regulation in cellular proliferation and differentiation. *Cell* 158, 1281–1292.

Goodarzi, H., Zhang, S., Buss, C.G., Fish, L., Tavazoie, S., and Tavazoie, S.F. (2014). Metastasis-suppressor transcript destabilization through TARBP2 binding of mRNA hairpins. *Nature* 513, 256–260.

Gromak, N., Dienstbier, M., Macias, S., Plass, M., Eyra, E., Cáceres, J.F., and Proudfoot, N.J. (2013). Drosha regulates gene expression independently of RNA cleavage function. *Cell Rep.* 5, 1499–1510.

Grosswendt, S., Filipchuk, A., Manzano, M., Klironomos, F., Schilling, M., Herzog, M., Gottwein, E., and Rajewsky, N. (2014). Unambiguous identification of miRNA:target site interactions by different types of ligation reactions. *Mol. Cell* 54, 1042–1054.

Gu, W., Shirayama, M., Conte, D., Jr., Vasale, J., Batista, P.J., Claycomb, J.M., Moresco, J.J., Youngman, E.M., Keys, J., Stoltz, M.J., et al. (2009). Distinct argonaute-mediated 22G-RNA pathways direct genome surveillance in the *C. elegans* germline. *Mol. Cell* 36, 231–244.

Gu, S., Jin, L., Zhang, Y., Huang, Y., Zhang, F., Valdmanis, P.N., and Kay, M.A. (2012). The loop position of shRNAs and pre-miRNAs is critical for the accuracy of dicer processing in vivo. *Cell* 151, 900–911.

Hafner, M., Landthaler, M., Burger, L., Khorshid, M., Hausser, J., Berninger, P., Rothballer, A., Ascano, M., Jr., Jungkamp, A.C., Munschauer, M., et al. (2010). Transcriptome-wide identification of RNA-binding protein and microRNA target sites by PAR-CLIP. *Cell* 141, 129–141.

- Han, J., Lee, Y., Yeom, K.H., Nam, J.W., Heo, I., Rhee, J.K., Sohn, S.Y., Cho, Y., Zhang, B.T., and Kim, V.N. (2006). Molecular basis for the recognition of primary microRNAs by the Drosha-DGCR8 complex. *Cell* 125, 887–901.
- Han, J., Pedersen, J.S., Kwon, S.C., Belair, C.D., Kim, Y.K., Yeom, K.H., Yang, W.Y., Haussler, D., Belloch, R., and Kim, V.N. (2009). Posttranscriptional crossregulation between Drosha and DGCR8. *Cell* 136, 75–84.
- Hellwig, S., and Bass, B.L. (2008). A starvation-induced noncoding RNA modulates expression of Dicer-regulated genes. *Proc. Natl. Acad. Sci. USA* 105, 12897–12902.
- Hutvagner, G., McLachlan, J., Pasquinelli, A.E., Bálint, E., Tuschl, T., and Zamore, P.D. (2001). A cellular function for the RNA-interference enzyme Dicer in the maturation of the let-7 small temporal RNA. *Science* 293, 834–838.
- Jungkamp, A.C., Stoeckius, M., Mecnas, D., Grün, D., Mastrobuoni, G., Kempa, S., and Rajewsky, N. (2011). In vivo and transcriptome-wide identification of RNA binding protein target sites. *Mol. Cell* 44, 828–840.
- Karginov, F.V., Cheloufi, S., Chong, M.M., Stark, A., Smith, A.D., and Hannon, G.J. (2010). Diverse endonucleolytic cleavage sites in the mammalian transcriptome depend upon microRNAs, Drosha, and additional nucleases. *Mol. Cell* 38, 781–788.
- Kim, V.N., Han, J., and Siomi, M.C. (2009). Biogenesis of small RNAs in animals. *Nat. Rev. Mol. Cell Biol.* 10, 126–139.
- Knight, S.W., and Bass, B.L. (2001). A role for the RNase III enzyme DCR-1 in RNA interference and germ line development in *Caenorhabditis elegans*. *Science* 293, 2269–2271.
- Kotaja, N., Bhattacharyya, S.N., Jaskiewicz, L., Kimmins, S., Parvinen, M., Filipowicz, W., and Sassone-Corsi, P. (2006). The chromatoid body of male germ cells: similarity with processing bodies and presence of Dicer and microRNA pathway components. *Proc. Natl. Acad. Sci. USA* 103, 2647–2652.
- Kozomara, A., and Griffiths-Jones, S. (2014). miRBase: annotating high confidence microRNAs using deep sequencing data. *Nucleic Acids Res.* 42 (Database issue), D68–D73.
- Lau, P.W., Guiley, K.Z., De, N., Potter, C.S., Carragher, B., and MacRae, I.J. (2012). The molecular architecture of human Dicer. *Nat. Struct. Mol. Biol.* 19, 436–440.
- Lebedeva, S., Jens, M., Theil, K., Schwanhäusser, B., Selbach, M., Landthaler, M., and Rajewsky, N. (2011). Transcriptome-wide analysis of regulatory interactions of the RNA-binding protein HuR. *Mol. Cell* 43, 340–352.
- Lee, Y., Ahn, C., Han, J., Choi, H., Kim, J., Yim, J., Lee, J., Provost, P., Rådmark, O., Kim, S., and Kim, V.N. (2003). The nuclear RNase III Drosha initiates microRNA processing. *Nature* 425, 415–419.
- Macias, S., Plass, M., Stajuda, A., Michlewski, G., Eyra, E., and Cáceres, J.F. (2012). DGCR8 HITS-CLIP reveals novel functions for the Microprocessor. *Nat. Struct. Mol. Biol.* 19, 760–766.
- MacRae, I.J., Zhou, K., and Doudna, J.A. (2007). Structural determinants of RNA recognition and cleavage by Dicer. *Nat. Struct. Mol. Biol.* 14, 934–940.
- Murchison, E.P., Stein, P., Xuan, Z., Pan, H., Zhang, M.Q., Schultz, R.M., and Hannon, G.J. (2007). Critical roles for Dicer in the female germline. *Genes Dev.* 21, 682–693.
- Nejepinska, J., Malik, R., Filkowski, J., Flemr, M., Filipowicz, W., and Svoboda, P. (2012). dsRNA expression in the mouse elicits RNAi in oocytes and low adenosine deamination in somatic cells. *Nucleic Acids Res.* 40, 399–413.
- Okamura, K., Hagen, J.W., Duan, H., Tyler, D.M., and Lai, E.C. (2007). The mirtron pathway generates microRNA-class regulatory RNAs in *Drosophila*. *Cell* 130, 89–100.
- Park, J.E., Heo, I., Tian, Y., Simanshu, D.K., Chang, H., Jee, D., Patel, D.J., and Kim, V.N. (2011). Dicer recognizes the 5' end of RNA for efficient and accurate processing. *Nature* 475, 201–205.
- Ricci, E.P., Kucukural, A., Cenik, C., Mercier, B.C., Singh, G., Heyer, E.E., Ashar-Patel, A., Peng, L., and Moore, M.J. (2014). Stauf1 senses overall transcript secondary structure to regulate translation. *Nat. Struct. Mol. Biol.* 21, 26–35.
- Ruby, J.G., Jan, C.H., and Bartel, D.P. (2007). Intronic microRNA precursors that bypass Drosha processing. *Nature* 448, 83–86.
- Sarkies, P., and Miska, E.A. (2013). RNAi pathways in the recognition of foreign RNA: antiviral responses and host-parasite interactions in nematodes. *Biochem. Soc. Trans.* 41, 876–880.
- Schisa, J.A., Pitt, J.N., and Priess, J.R. (2001). Analysis of RNA associated with P granules in germ cells of *C. elegans* adults. *Development* 128, 1287–1298.
- Seila, A.C., Calabrese, J.M., Levine, S.S., Yeo, G.W., Rahl, P.B., Flynn, R.A., Young, R.A., and Sharp, P.A. (2008). Divergent transcription from active promoters. *Science* 322, 1849–1851.
- Shi, Z., Montgomery, T.A., Qi, Y., and Ruvkun, G. (2013). High-throughput sequencing reveals extraordinary fluidity of miRNA, piRNA, and siRNA pathways in nematodes. *Genome Res.* 23, 497–508.
- Sinkkonen, L., Hugenschmidt, T., Filipowicz, W., and Svoboda, P. (2010). Dicer is associated with ribosomal DNA chromatin in mammalian cells. *PLoS ONE* 5, e12175.
- Taylor, D.W., Ma, E., Shigematsu, H., Cianfrocco, M.A., Noland, C.L., Nagayama, K., Nogales, E., Doudna, J.A., and Wang, H.W. (2013). Substrate-specific structural rearrangements of human Dicer. *Nat. Struct. Mol. Biol.* 20, 662–670.
- Tian, Y., Simanshu, D.K., Ma, J.B., Park, J.E., Heo, I., Kim, V.N., and Patel, D.J. (2014). A phosphate-binding pocket within the platform-PAZ-connector helix cassette of human Dicer. *Mol. Cell* 53, 606–616.
- Updike, D.L., Hachey, S.J., Kreher, J., and Strome, S. (2011). P granules extend the nuclear pore complex environment in the *C. elegans* germ line. *J. Cell Biol.* 192, 939–948.
- Valen, E., Preker, P., Andersen, P.R., Zhao, X., Chen, Y., Ender, C., Dueck, A., Meister, G., Sandelin, A., and Jensen, T.H. (2011). Biogenic mechanisms and utilization of small RNAs derived from human protein-coding genes. *Nat. Struct. Mol. Biol.* 18, 1075–1082.
- Wan, Y., Qu, K., Zhang, Q.C., Flynn, R.A., Manor, O., Ouyang, Z., Zhang, J., Spitale, R.C., Snyder, M.P., Segal, E., and Chang, H.Y. (2014). Landscape and variation of RNA secondary structure across the human transcriptome. *Nature* 505, 706–709.
- Welker, N.C., Maity, T.S., Ye, X., Aruscavage, P.J., Krauchuk, A.A., Liu, Q., and Bass, B.L. (2011). Dicer's helicase domain discriminates dsRNA termini to promote an altered reaction mode. *Mol. Cell* 41, 589–599.
- Zamudio, J.R., Kelly, T.J., and Sharp, P.A. (2014). Argonaute-bound small RNAs from promoter-proximal RNA polymerase II. *Cell* 156, 920–934.
- Zhang, H., Kolb, F.A., Brondani, V., Billy, E., and Filipowicz, W. (2002). Human Dicer preferentially cleaves dsRNAs at their termini without a requirement for ATP. *EMBO J.* 21, 5875–5885.

Unraveling the Biology of a Fungal Meningitis Pathogen Using Chemical Genetics

Jessica C.S. Brown,^{1,6} Justin Nelson,² Benjamin VanderSluis,² Raamesh Deshpande,² Arielle Butts,³ Sarah Kagan,⁴ Itzhack Polacheck,⁴ Damian J. Krysan,^{3,5} Chad L. Myers,^{2,*} and Hiten D. Madhani^{1,*}

¹Department of Biochemistry and Biophysics, University of California, San Francisco, San Francisco, CA 94158, USA

²Department of Computer Science and Engineering, University of Minnesota, Minneapolis, MN 55455, USA

³Department of Chemistry, University of Rochester Medical Center, Rochester, NY 14643, USA

⁴Department of Clinical Microbiology and Infection Diseases, Hadassah-Hebrew University Medical Center, Jerusalem 91120, Israel

⁵Departments of Pediatrics and Microbiology/Immunology, University of Rochester School of Medicine and Dentistry, Rochester, NY 14643, USA

⁶Present address: Department of Pathology, Division of Microbiology and Immunology, University of Utah School of Medicine, Salt Lake City, UT 84112, USA

*Correspondence: cmyers@cs.umn.edu (C.L.M.), hitenmadhani@gmail.com (H.D.M.)

<http://dx.doi.org/10.1016/j.cell.2014.10.044>

SUMMARY

The fungal meningitis pathogen *Cryptococcus neoformans* is a central driver of mortality in HIV/AIDS. We report a genome-scale chemical genetic data map for this pathogen that quantifies the impact of 439 small-molecule challenges on 1,448 gene knock-outs. We identified chemical phenotypes for 83% of mutants screened and at least one genetic response for each compound. *C. neoformans* chemical-genetic responses are largely distinct from orthologous published profiles of *Saccharomyces cerevisiae*, demonstrating the importance of pathogen-centered studies. We used the chemical-genetic matrix to predict novel pathogenicity genes, infer compound mode of action, and to develop an algorithm, O2M, that predicts antifungal synergies. These predictions were experimentally validated, thereby identifying virulence genes, a molecule that triggers G2/M arrest and inhibits the Cdc25 phosphatase, and many compounds that synergize with the antifungal drug fluconazole. Our work establishes a chemical-genetic foundation for approaching an infection responsible for greater than one-third of AIDS-related deaths.

INTRODUCTION

Invasive fungal infections are notoriously difficult to diagnose and treat, resulting in high mortality rates, even with state-of-the-art treatments. The three most common pathogenic agents are *Cryptococcus neoformans*, *Candida albicans*, and *Aspergillus fumigatus* (Mandell et al., 2010). These organisms are opportunistic fungi that prey on individuals with varying degrees of immune deficiency. Susceptible patient populations include premature infants, diabetics, individuals with liver disease, chemotherapy patients, organ transplant recipients, and those

infected with HIV (Mandell et al., 2010). Compounding the clinical challenge is the slow pace of antifungal drug development: only a single new class of drugs (the echinocandins) has been approved for use in the United States in the last 30 years (Butts and Krysan, 2012; Mandell et al., 2010; Roemer et al., 2011).

Fungal infections are estimated to cause 50% of deaths related to AIDS and have been termed a “neglected epidemic” (Armstrong-James et al., 2014). The fungus chiefly responsible for deaths in this population is *C. neoformans* (Armstrong-James et al., 2014). *C. neoformans* is an encapsulated basidiomycetous haploid yeast distantly related to *Saccharomyces cerevisiae* and *Schizosaccharomyces pombe*. A 2009 CDC study estimated that ~1 million infections and ~600,000 deaths annually are caused by *C. neoformans*, exceeding the estimated worldwide death toll from breast cancer (Lozano et al., 2012; Park et al., 2009). *C. neoformans* is widespread in the environment and exposure occurs through inhalation of desiccated yeast or spores (Heitman et al., 2011). In immunocompromised patients, *C. neoformans* replicates and disseminates, causing meningo-encephalitis that is lethal without treatment (Heitman et al., 2011). Induction therapy involves flucytosine and intravenous infusions of amphotericin B (Loyse et al., 2013). Both drugs are highly toxic, difficult to administer, and neither is readily available in the areas with the highest rates of disease. The current recommendation for Cryptococcosis treatment is at least a year of therapy, which is difficult to accomplish in resource-limited settings (WHO, 2011). Thus, as is the case with infections caused by other fungal pathogens, effective treatment of cryptococcal infections is limited by the efficacy, toxicity, and availability of current pharmaceuticals.

We implemented chemogenomic profiling to approach the challenges of therapeutic development in *C. neoformans*. This method involves the systematic measurement of the impact of compounds on the growth of defined null mutants to produce a chemical-genetic map. Such a map represents a quantitative description composed of numerical scores indicative of the growth behavior of each knockout mutant under each chemical condition. Cluster analysis of the growth scores for large

numbers of mutants under many chemical conditions can reveal genes that function in the same pathway and even those whose products are part of the same protein complex (Collins et al., 2007; Parsons et al., 2004; Parsons et al., 2006). In addition, the identity of genes whose mutation produce resistance or sensitivity is useful for uncovering compound mode of action (MOA) (Hillenmeyer et al., 2008; Jiang et al., 2008; Nichols et al., 2011; Parsons et al., 2006; Xu et al., 2007; Xu et al., 2009). Large-scale studies have been restricted to model organisms for which gene deletion collections have been constructed, namely *S. cerevisiae*, *S. pombe*, and *Escherichia coli* K12 (Hillenmeyer et al., 2008; Nichols et al., 2011; Parsons et al., 2006). However, as none of these are pathogens, the extent to which the resulting insights translate to pathogenic organisms is unknown. A variation on chemogenomic profiling, chemically-induced haploinsufficiency, was first developed using a diploid heterozygote gene deletion library *S. cerevisiae* to identify compound MOA. This method, which identifies genes that impact compound sensitivity based on a two-fold gene dosage change, is suited for diploid organisms and has been used in the pathogen *C. albicans* (Jiang et al., 2008; Xu et al., 2007; Xu et al., 2009).

We report here the generation of a large-scale chemogenomic map for *C. neoformans* using defined, commonly available knockout mutants, assessments of data quality, and extensive experimental verification. Comparisons of the *C. neoformans* profile with two large-scale published profiles from *S. cerevisiae* revealed that for most types of compounds, the chemical-genetic interactions are distinct even among orthologous genes, emphasizing the importance of pathogen-focused investigation. We used nearest-neighbor analysis to predict new genes involved in polysaccharide capsule formation and infectivity, which we validated through experiment. We also utilized genetic responses to predict the G2/M phase of the cell cycle and the Cdc25 phosphatase as targets of a thiazolidone-2,4-dione derivative, which we confirmed in vivo and in vitro. Finally, because of the unmet need for improved antifungal drug efficacy, we developed a new algorithm, O2M, to predict new compound synergies based on the profiles of pairs known to be synergistic. Experimental tests demonstrate that the method performs vastly better than random expectation, thereby enabling the identification of synergistic compound combinations. Our studies establish a chemical-genetic foundation to approach the biology and treatments of *C. neoformans* infections, which are responsible for more than one-third of HIV/AIDS deaths worldwide.

RESULTS

A Chemical-Genetic Map of *C. neoformans*

We assembled 1,448 *C. neoformans* gene deletion strains (Chun et al., 2011; Liu et al., 2008) (Table S1 available online), corresponding to a substantial fraction of 6,967 predicted *C. neoformans* genes (Janbon et al., 2014), and a collection of compounds for screening (Table 1). Compounds were selected based on cost and literature evidence that they could inhibit the growth of fungi. Where feasible, compounds were chosen that are known to target specific biological processes. For

each small molecule, we determined an approximate minimum inhibitory concentration (MIC) in agar, then measured growth of the knockout collection on each small molecule at 50%, 25%, and 12.5% MIC using high density agar plate colony arrays and a robotic replicator. We then measured the size of each colony using flatbed scanning and colony measurement software (Dittmar et al., 2010). We performed a minimum of four replicate colony measurements for each mutant-condition pair. Plate-based assays are subject to known nonbiological effects, such as spatial patterns. To mitigate these errors, a series of corrective measures were implemented using approaches described previously, including manual filtration of noisy data, spatial effect normalization and machine learning-based batch correction (Baryshnikova et al., 2010). In addition, the data for each deletion mutant and compound was centered and normalized. Each mutant-small molecule combination was assigned a score with positive scores representing relative resistance and negative scores representing compound sensitivity (Table S2). A global summary of the processed data organized by hierarchical clustering is shown in Figure 1A.

The importance and validity of the computational corrections is shown in Figures 1B and S1. We estimated how reproducible the chemical-genetic profiles were by calculating the correlation scores for data obtained for different concentrations of the same small molecule (purple). This measures the degree of overlap between the overall chemical-genetic profiles, which are themselves each composed of a score for each mutant-small molecule combination. We found significant correlation ($p = 2.67 \times 10^{-176}$) between data obtained for different concentrations of the same small molecule compared to those between profiles generated by data set randomization, suggesting significant reproducibility. Moreover, correlation scores between chemical-genetic profiles of different concentrations of different compounds (gray) are centered at approximately 0 (Figure 1B). This difference in correlation scores is apparent even when comparing experiments performed on the same day, when spurious batch signal can contribute to false positives (Baryshnikova et al., 2010). Our batch-correction algorithms resulted in same-batch screening data with strong positive correlation scores for the same compounds but correlation scores close to zero for different compounds (Figure S1), demonstrating successful removal of spurious signal (Baryshnikova et al., 2010). We compared chemical-genetic profiles between compounds in the azole family (Figure 1C). Despite the fact that the azoles tested include those of diverse uses, from agricultural pesticides to FDA-approved drugs (Table 1), many exhibit a significant profile correlation ($p = 2.82 \times 10^{-6}$), further indicating significant signal in the data. As a final assessment, we performed hypergeometric testing across all compounds to determine whether the same sensitive gene knockouts (defined by $Z < -2.5$) are identified at different concentrations of the same compounds. Using a Bonferroni-corrected p value cutoff, nearly all compounds display significant overlap of responsive genes at different concentrations (Figure 1D).

We assigned at least one phenotype (sensitivity or resistance to a compound) to 1,198 of 1,448 mutants (Figure 1E, Tables S2, S3, and S4). Of these, 855 exhibit one to ten phenotypes, while remaining 343 displayed from 11 to 146 phenotypes. Gene

Table 1. Small Molecules and Targets

Inhibitor (Activator)	Highest Screening Conc.	Process/Enzyme	Category	Pubchem ID	FDA Approval?
1-10 phenanthroline hydrochloride monohydrate	2 uM	broad/transition metal complexes	broad spectrum	2723715	no
2-aminobenzothiazole	30 uM	cytoskeleton function/ kinesin Kip1	cell structure	8706	no
2-hydroxyethylhydrazine	0.156%	lipid synthesis/phospholipid methylation	lipid biosynthesis	8017	no
3-aminotriazole	6.25 mM	histidine synthesis/IMP dehydratase	metabolism	1639	no
4-hydroxytamoxifen	1.56 uM	estrogen receptor (mammals)	signaling	449459	yes
5-fluorocytosine	2.5 ug/ml	DNA/RNA biosynthesis	DNA homeostasis/ protein synthesis	3366	yes
5-methyltryptophan	8 mM	tryptophan synthesis	metabolism	150990	no
Abietic acid	1 mM	lipid synthesis/lipoxygenase	lipid biosynthesis	10569	no
Acifluorfen methyl	156.25 ng/ml	porphyrin synthesis/ protoporphyrinogen oxidase	metabolism	91642	no
(Aconitine)	200 ug/ml	membrane potential/Na ⁺ channels (mammals)	membrane polarization	245005	no
Aflatoxin B1	100 ug/ml	DNA damaging agent	DNA homeostasis	14403	no
Agelastine D	5 ug/ml	membrane potential/Na ⁺ / K ⁺ -ATPase (mammals)	membrane polarization	46231918	no
Alamethicin/U-22324	60 uM	membrane integrity/forms a voltage-dependent ion channel	membrane polarization	16132042	no
Alexidine dihydrochloride	125 ug/ml	antimicrobial/mitochondria	mitochondria	102678	yes
Allantoin	100 ug/ml	nitrogen-rich compound	metabolism	204	topical
Alternariol	2.5 ug/ml	cholinesterase inhibitor/sodium channel activator and DNA supercoiling/topoisomerase I	broad spectrum	5359485	no
Aluminum sulfate	1.5625 mM	unknown	unknown	24850	no
(Amantadine hydrochloride)	1.25 mM	neurotransmitter release/ glutamate receptor	signaling	64150	yes
Amiodarone	60 ug/ml	membrane potential/Na ⁺ / K ⁺ -ATPase (mammals)	membrane polarization	2157	yes
(Ammonium persulfate)	50 mM	reactive oxygen species	apoptosis/stress response/ damage response	62648	no
Amphotericin B	1 ug/ml	lipid biosynthesis/ergosterol	membrane integrity	5280965	yes
Andrastin A	4 ug/ml	protein modification/ farnesyltransferase	protein trafficking	6712564	no
Anisomycin	50 uM	translation/peptidyl transferase	gene expression	253602	no
Antimycin	100 ug/ml	respiration/cytochrome B	metabolism	14957	no
Apicidin	312.5 ng/ml	chromatin regulation/HDACs	gene expression	6918328	no
Artemisinin	312.5 mM	iron metabolism/hematin detoxification	metabolism	68827	yes
Ascomycin	3.125 uM	signaling/calcineurin	signaling	6437370	yes
Azide	62.5 uM	respiration/cytochrome C oxidase	metabolism	33558	no
Barium chloride	16 mM	metal homeostasis/diverse	broad spectrum/unknown	25204	no
Bafilomycin	4 ug/ml	autophagy/vacuolar-type H ⁺ -ATPase	protein turnover	6436223	no
Bathocuproine disulphonic acid (BCS)	3 mM	copper acquisition	metabolism	16211287	no

(Continued on next page)

Table 1. Continued

Inhibitor (Activator)	Highest Screening Conc.	Process/Enzyme	Category	Pubchem ID	FDA Approval?
Bathophenanthroline disulfonate (BPS)	300 uM	iron acquisition/Fet3-Ftr1	metabolism	65368	no
Benomyl	100 ug/ml	cytoskeleton function/tubulin	cell structure	28780	no
(Betulinic acid)	64 ug/ml	protein degradation/proteasome	protein turnover	64971	no
Bifonazole	50 ug/ml	lipid biosynthesis/HMG-CoA and ergosterol biosynthesis	membrane integrity	2378	no
Brefeldin A	40 ug/ml	ER-Golgi Transport/ARF GEF	secretion	5287620	no
Calcium chloride	16 mM	metal homeostasis/diverse	broad spectrum	5284359	no
Caffeine	2.5 mM	DNA damage checkpoint/ATM	DNA homeostasis	2519	no
Calcium ionophore A23187	2.5 ug/ml	membrane integrity/peptide that acts as ionophore	membrane integrity	40486	no
Calcofluor white	500 ug/ml	cell wall synthesis/chitin and cellulose	cell wall	6108780	no
Camptothecin	500 ug/ml	DNA supercoiling/topoisomerase I	DNA homeostasis	24360	analog
Castanospermine	2.4 mM	protein modification/glycosidation	protein modification	54445	derivative
Cadmium chloride	1 mM	metal homeostasis/diverse	broad spectrum/unknown	24947	no
Cerulenin	312.5 ng/ml	fatty acid synthesis/beta-ketoacyl-acyl carrier protein synthase	lipid biosynthesis	5282054	no
Cesium chloride	128 mM	metal homeostasis/diverse	broad spectrum/unknown	24293	no
Chlorpromazine hydrochloride	1.5625 uM	phenothiazine antipsychotic drug (mammals)/dopamine, serotonin, and other neuroreceptors	signaling	6240	yes
Chromium (III) chloride	8 mM	metal homeostasis/diverse	broad spectrum	16211596	no
Ciclopirox olamine	750 ng/ml	iron acquisition and other	metabolism	38911	yes
Cisplatin	100 ug/ml	DNA synthesis	DNA homeostasis	157432	yes
Climbazole	0.03125%	lipid biosynthesis/ergosterol biosynthesis and respiration/cytochrome P450	broad spectrum	37907	topical
Clotrimazole	500 nM	lipid biosynthesis/ergosterol biosynthesis	membrane integrity	2812	yes
Colistin	1 mg/ml	membrane integrity	membrane integrity	5311054	yes
Congo red	0.0625%	cell wall synthesis/chitin, cellulose, and glucan	cell wall	11313	no
Coniine	0.15625%	neurosignaling (mammals)/nicotinic receptor	signaling	441072	no
(Crystal violet)	0.0012500%	oxidative stress inducer	stress response	11057	topical
CuCl ₂	8 mM	copper homeostasis/diverse	metabolism	24014	no
Cycloheximide	1.875 ug/ml	translation/ribosome	gene expression	6197	no
Cyclopiazonic acid	15.625 uM	ion transport and cell polarization (mammals)/Ca ²⁺ -ATPase	metabolism	54682463	no
Cyclosporin	75 ug/ml	signaling/calcineurin	signaling	5284373	yes
Cyproconazole	1.5625 ug/ml	lipid biosynthesis/ergosterol biosynthesis	membrane integrity	86132	no
Cyprodinil	10 ug/ml	methionine biosynthesis	metabolism	86367	no
Daphnetin	100 uM	signaling/PKA, PKC, EGR receptor, others	signaling	5280569	no
Desipramine hydrochloride	250 uM	neurosignaling (mammals)/norepinephrine transporter	signaling	65327	yes

(Continued on next page)

Table 1. Continued

Inhibitor (Activator)	Highest Screening Conc.	Process/Enzyme	Category	Pubchem ID	FDA Approval?
Dyclonine hydrochloride	3.125 uM	lipid biosynthesis/ergosterol biosynthesis	membrane integrity	68304	yes
Emetine dihydrochloride hydrate	5 mM	translation/ribosome	gene expression	3068143	yes
Emodin	62.5 uM	signaling/CK2, others	signaling	3220	yes
Erlotinib	50 ug/ml	signaling (mammals)/EGFR tyrosine kinase	signaling	176870	yes
FeCl ₃	32 mM	iron acquisition, metal homeostasis	metabolism	24380	no
Fenoxanil	80 ug/ml	melanin biosynthesis	metabolism	11262655	no
Fenpropimorph	2.5 ug/ml	sterol synthesis	lipid biosynthesis	93365	no
FK506	312.5 ng/ml	signaling/calcineurin	signaling	445643	yes
Fluconazole	10 ug/ml	lipid biosynthesis/ergosterol biosynthesis	membrane integrity	3365	yes
Fluspirilene	25 uM	antipsychotic drug, mechanism of action unknown	unknown	3396	yes
Gallium (III) nitrate	25 mM	metal homeostasis/diverse	broad spectrum	57352728	no
Geldanamycin	2 uM	protein folding/Hsp90	protein folding	5288382	trials
(H ₂ O ₂)	6 mM	reactive oxygen species	apoptosis/ stress response/damage response	784	topical
Haloperidol	125 uM	phenothiazine antipsychotic drug (mammals)/dopamine, serotonin, and other neuroreceptors	signaling	3559	yes
Harmine hydrochloride	1 mM	cell differentiation (mammals)/PPARgamma	signaling	5359389	yes
Hydroxyurea	12.5 mM	DNA replication/replication fork progression	DNA homeostasis	3657	yes
Hygromycin	37.5 ug/ml	translation/ribosome	gene expression	35766	no
Imazalil	25 ug/ml	lipid biosynthesis/ergosterol synthesis	membrane integrity	37175	no
Iodoacetate	500 uM	protein degradation/cysteine peptidases	protein turnover	5240	no
Itraconazole	1.5625 ug/ml	lipid biosynthesis/ergosterol synthesis	membrane integrity	55283	yes
K252a	10 ug/ml	signaling/variety of kinases	signaling	127357	trials
Latrunculin	25 uM	cytoskeleton function/actin	cell structure	445420	no
Lead (II) nitrate	64 mM	metal homeostasis/diverse	broad spectrum	24924	no
Leptomycin	1.25 ug/ml	nucleocytoplasmic transport/Crm1	gene expression	6917907	no
LiCl	37.5 mM	metal homeostasis/diverse	broad spectrum/ unknown	433294	no
Lovastatin	37.5 ug/ml	sterol synthesis/HMG CoA reductase	metabolism	53232	yes
LY 294002	375 uM	signaling/PI3K	signaling	3973	no
Magnesium chloride	150 mM	metal homeostasis/diverse	broad spectrum	21225507	no
Malachite green	3.125 ug/ml	antimicrobial/unknown	antimicrobial	11294	no
Manganese sulfate	128 mM	metal homeostasis/diverse	metabolism	177577	no
Mastoparan	5 uM	signaling/G-proteins	signaling	5464497	no
(Menadione)	150 uM	vitamin K3/reactive oxygen species	diverse	4055	yes

(Continued on next page)

Table 1. Continued

Inhibitor (Activator)	Highest Screening Conc.	Process/Enzyme	Category	Pubchem ID	FDA Approval?
Menthol	1 mM	voltage-dependent ion channels (mammals)/sodium channel	signaling	16666	yes
Methotrexate	2.5 uM	folate synthesis/DHFR	metabolism	126941	yes
Methyl methanesulfonate (MMS)	0.0165%	DNA replication/replication fork progression	DNA homeostasis	4156	no
Methylbenzethonium chloride (MBT)	0.25%	antimicrobial	antimicrobial	5702238	topical
MG132	12.5 uM	protein degradation/proteasome	protein turnover	462382	no
Miconazole	6.25 ug/ml	lipid biosynthesis/ergosterol synthesis	membrane integrity	4189	yes
Mitomycin C	12 uM	DNA damaging agent	DNA homeostasis	5746	yes
Myclobutanil	2 ug/ml	lipid biosynthesis/ergosterol synthesis	membrane integrity	6336	no
Mycophenolic acid	2.5 ug/ml	GMP synthesis/IMP dehydrogenase	metabolism	446541	yes
Myriocin	12.5 ug/ml	sphingolipid synthesis	metabolism	6438394	analog
NA8		unknown	unknown		no
(NaCl)	37.5 mM	osmotic regulation/HOG pathway	stress response	5234	yes
(NaNO ₂)	150 uM	reactive nitrogen species	stress response	23668193	no
Neomycin sulfate	2.4 mM	protein synthesis/ribosome	gene expression	8378	yes
Nicotinamide	25 uM	chromatin regulation/sirtuins	gene expression	936	yes
Nigericin	100 ug/ml	membrane integrity/ion gradient	membrane polarization	34230	no
Nikkomycin	5 ug/ml	chitin synthesis	cell wall	72479	trials
NiSO ₄	1 mM	antifungal/diverse	antifungal	5284429	no
Nocodazole	30 uM	cytoskeleton function/tubulin	cell structure	4122	no
Ophiobolin A	62.5 ng/ml	signaling/calmodulin	signaling	5281387	no
Parthenolide	150 uM	immune and inflammatory response/NF-kB	signaling	6473881	no
Pentamidine isethionate	500 uM	antimicrobial/mitochondrial function	antimicrobial	8813	yes
pH	8.0, 8.5, 9.0	pH homeostasis	diverse		no
Phenylarsine oxide	2.5 uM	broad/XCXXCX protein phosphatases	broad spectrum	4778	no
Picoxystrobin	6.25 ug/ml	quinone outside inhibitor class/fungal cytochrome bcl	mitochondria	11285653	no
(Plumbagin)	2.8 uM	reactive oxygen species	stress response	10205	no
PMSF	10 mM	vacuolar proteolysis/proteinase B	signaling	4784	no
Polyoxin B	200 ug/ml	chitin synthesis	cell wall	3084093	no
Povidone iodine	2%	antimicrobial	antimicrobial	410087	topical
Prussian blue	75 mM	monocation chelator	metabolism	16211064	yes
Quinic acid	2 mM	antimicrobial	antimicrobial	6508	no
Rapamycin	0.125 uM	signaling/TOR kinases	signaling	5284616	yes
Rubidium chloride	150 mM	potassium metabolism/competitor	metabolism	62683	no
Rifamycin SV monosodium salt	200 ug/ml	RNA synthesis/RNA polymerase	gene expression	6324616	yes
S10		unknown	unknown		no
S8		unknown	unknown		no

(Continued on next page)

Table 1. Continued

Inhibitor (Activator)	Highest Screening Conc.	Process/Enzyme	Category	Pubchem ID	FDA Approval?
S-aminoethyl-L-cysteine (thialysine)	10 uM	amino acid metabolism/lysine analog	metabolism	20048	no
SDS	0.0015625%	cell membrane integrity	membrane integrity	3423265	no
Selumetinib	150 ug/ml	signaling/MAPK (ERK)	signaling	10127622	trials
Sertraline	15 ug/ml	neurosignaling (mammals)/serotonin reuptake	neurosignaling	68617	yes
Sodium azide	62.5 uM	respiration/cytochrome oxidase	mitochondria	33557	no
Sodium borate	10 mM	antimicrobial/diverse	antimicrobial	21749317	no
Sodium hydrosulfite	6.25 mM	antimicrobial, counteracts some antimicrobials	antimicrobial	24489	no
Sodium iodide	75 mM	antimicrobial	antimicrobial	5238	yes
Sodium metavanadate	10 mM	signaling/protein phosphotyrosine phosphatases	signaling	4148882	no
(Sodium molybdate)	64 mM	respiration/oxygen uptake	diverse	61424	no
Sodium selenite	4 mM	respiration/oxygen uptake	diverse	16210997	yes
Sodium sulfite	100 mM	ATP synthesis and accumulation/unknown	metabolism	24437	no
Sodium tungstate	64 mM	metal homeostasis/diverse	broad spectrum/unknown	150191	no
Sorafenib	100 uM	signaling/VEGF tyrosine kinase	signaling	216239	yes
Staurosporine	3 uM	signaling/PKC1	signaling	5279	yes
(STF-62247)	400 uM	autophagy	protein turnover	704473	trials
Sulfometuron methyl	100 ug/ml	branch chain amino acid synthesis/acetolactate synthase	metabolism	52997	no
Suloctidil	400 uM	Ca ²⁺ homeostasis in blood vessels (mammals)/putative Ca ²⁺ channel blocker	vascular system/metabolism	5354	formerly
Tamoxifen citrate	10 uM	estrogen signaling (mammals)/estrogen receptor, mixed agonist/antagonist	signaling	2733525	yes
Taurolidine	0.01%	antimicrobial/lipopolysaccharide detection and signaling	host defense	29566	yes
Tautomycin	250 nM	signaling/PP2A	signaling	3034761	no
Tellurite	0.1%	sulfate assimilation	metabolism	115037	no
Terbinafine	75 uM	sterol synthesis/squalene epoxidase	metabolism	1549008	yes
Thiabendazole	200 ug/ml	respiration/NADH oxidase	mitochondria	5430	yes
Thonzonium bromide	25 uM	antimicrobial, pH homeostasis/V-ATPase	broad spectrum	11102	yes
Tomatine	5 ug/ml	glycoalkaloid antifungal of unknown mechanism/ergosterol biosynthesis	antifungal/membrane integrity	28523	no
Trichostatin A	100 uM	chromatin regulation/HDACs	gene expression	444732	no
Trifluoperazine	200 uM	signaling/calmodulin	signaling	5566	yes
Trimethoprim	1.6 mg/ml	folate synthesis/DHFR	metabolism	5578	yes
Tunicamycin	2.5 ug/ml	glycosylation/Alg7	secretion	11104835	no
Usnic acid	25 ug/ml	antimicrobial	antimicrobial	6433557	trials
Valinomycin	20 uM	membrane integrity/potassium exclusion	membrane polarization	5649	no
Verrucarin	5 uM	protein biosynthesis/polysome	protein turnover	6437060	no
ZnCl ₂	4 mM	metal homeostasis/diverse	diverse	5727	no

A list of compounds used in this study, their targets, and the screening concentration.

deletions with the greatest number of phenotypes are *cnag_07622Δ* (encoding the COP9 signalosome subunit 1) and *cnag_05748Δ* (encoding a Nto1 subunit of the NuA3 histone acetyltransferase). Compounds that elicit the greatest number of responsive gene deletions (Figure 1F) are the heavy metal salt sodium tungstate and the trichothecene protein synthesis inhibitor verrucaric acid (Table S5), presumably reflecting the pleiotropic impact of these molecules on cells.

Gene Ontology Analysis Reveals Processes Associated with Drug Sensitivity

Drug influx and efflux is thought to be a major general determinant of microbial drug susceptibility (Fernández and Hancock, 2012), but we also sought functions involved in drug sensitivity. We investigated this question in an unbiased fashion by analyzing chemogenomic profiles using Gene Ontology (GO), a gene annotation approach useful for comparative analyses. We first identified annotated orthologs of *C. neoformans* genes represented in the deletion library and associated GO terms with these orthologs. We then determined whether the sensitive gene knockouts that respond to each small molecule are enriched for association with particular GO terms relative to a randomized control set (Figure 2, Table S6). We observed that protein transport-related terms are highly enriched, as are processes related to ubiquitin modification/proteolysis and vesicle-mediated transport. These terms are associated with nine and five compounds, respectively, suggesting that intracellular transport and ubiquitin-mediated protein turnover may play important general roles in drug sensitivity.

Comparison with *S. cerevisiae* Chemogenomic Profiling Data Sets

Chemogenomic profiling has been performed extensively in *S. cerevisiae*, allowing us to ask whether genetic responses to compounds were conserved. We performed a three-way comparison with two large-scale studies (Hillenmeyer et al., 2008; Parsons et al., 2006) (Figure 3A). Our data set has 46 compounds in common with Parsons et al. (2006) and 29 with Hillenmeyer et al. (2008); the two *S. cerevisiae* data sets had 15 compounds in common. First we identified genes whose knockouts exhibited a significant ($Z \leq -2.5$ or $\geq +2.5$) score ("responding") when treated with a small molecule used in more than one data set, then identified which of those genes had orthologs in both *S. cerevisiae* and *C. neoformans*. We then calculated how many orthologs responded in both data sets. To adjust for a greater starting number of common genes when comparing the *S. cerevisiae* data sets to each other and control for functional biases, we limited this comparison to genes that also have orthologs in the *C. neoformans* knockout collection. The blue labels for compounds in Figures 3B–3D indicate statistically significant similarities ($p \leq 0.05$) in drug responses. Nearly all of the compounds in common between the two *S. cerevisiae* studies display statistically significant overlap in the genes that produced sensitivity to a given compound, despite the very different experimental platforms that were used to assess drug sensitivity/resistance (13/15 cases; Figure 3B). In striking contrast, few compounds show significantly conserved genetic responses when comparing either *S. cerevisiae* data set with

the *C. neoformans* data. For the two *C. neoformans*–*S. cerevisiae* comparisons, only two of 46 compounds (Figure 3C) and one of 29 compounds (Figure 3D) show conserved responses, respectively.

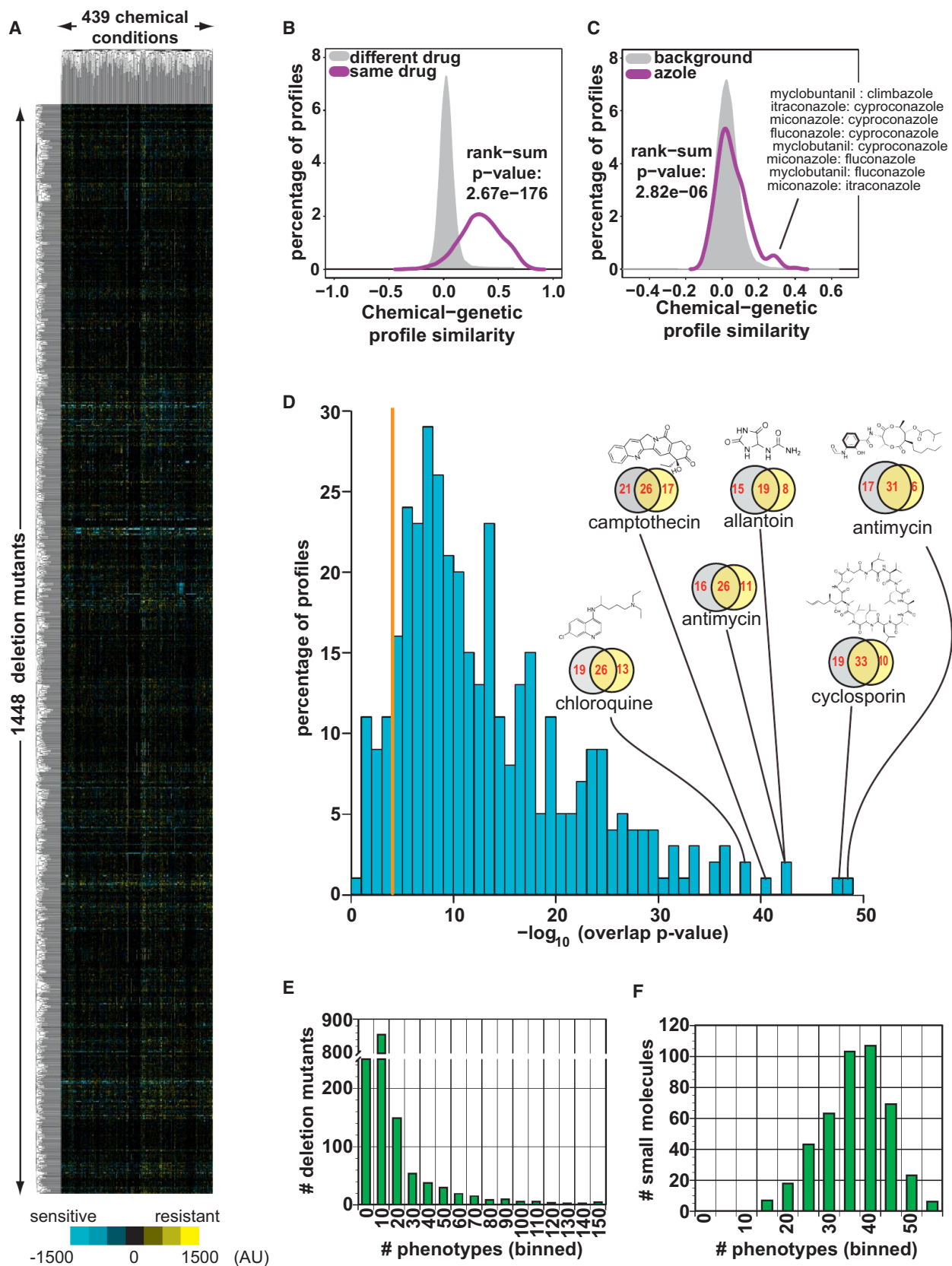
The responses to azole compounds exhibit limited response conservation between species. Comparing our data set with Parsons et al., the responses to fluconazole (FLC) and clotrimazole, the azoles in both data sets, do not show significant overlap (Figure 3C). Likewise, between our data set and Hillenmeyer et al. (2008), no gene orthologs respond to miconazole and clotrimazole in both data sets (Figure 3D). In contrast, between the two *S. cerevisiae* data sets, the only shared azole, clotrimazole, shows a significantly similar response (Figure 3B). We compared published work that examined the transcriptome responses of *S. cerevisiae* (Kuo et al., 2010) and *C. neoformans* (Florio et al., 2011) to FLC. We found that, while there was significant overlap in orthologous genes impacted in the two species, ($p = 1.6 \times 10^{-3}$), there were also considerable differences: 67% of the genes with an altered response in *C. neoformans* whose orthologs in *S. cerevisiae* did not exhibit significant change, (Table S7) (Kuo et al., 2010).

Using Chemical-Genetic Signatures to Identify Capsule Biosynthesis Mutants

Studies in *S. cerevisiae* have shown that the phenotypic signatures of gene deletions for genes that act in the same process or protein complex tend to be similar (Collins et al., 2007; Costanzo et al., 2010; Nichols et al., 2011; Parsons et al., 2004; Parsons et al., 2006). We reasoned that this property of could be used in a pathogen to identify candidates for new genes involved in virulence by simply testing gene deletions that displayed phenotypic profiles similar to those corresponding to known virulence factors.

C. neoformans harbors an inducible polysaccharide capsule that is unusual among fungi (Del Poeta, 2004; Doering, 2009; Haynes et al., 2011; Kumar et al., 2011; O'Meara and Alspaugh, 2012; O'Meara et al., 2010; Vecchiarelli et al., 2013). The principal polysaccharide component, glucuronylxylomannan (GXM), consists of a repeating glycan unit that has α -1,3-linked mannose backbone with side chains of β -linked glucuronic acid and xylose (Kozel et al., 2003). Capsule production is critical for virulence and the ability of *C. neoformans* to evade detection and destruction by the host immune system (Vecchiarelli et al., 2013).

To identify candidates for genes involved in capsule formation and/or attachment, we organized our data set using hierarchical clustering of growth phenotypes produced by compound exposure. We focused on two clusters, each containing a gene(s) previously implicated in capsule biosynthesis: *PBX1* and *CPL1* (Liu et al., 2008; Liu et al., 2007b) in one cluster (Figure 4A) and *CAP60* (Chang and Kwon-Chung, 1998) in a second cluster (Figure 4B). The *pbx1Δ/cpl1Δ* cluster contains nine genes and the *cap60Δ* cluster seven. We quantified capsule accumulation after induction by computing the ratio of the diameter of the cell and capsule to the diameter of the cell alone (Figures 4C and 4D). Wild-type cells exhibit high capsule production, *pbx1Δ* mutants display a partial defect (Liu et al., 2007a) and *cpl1Δ* and *cap60Δ* mutants are acapsular (Chang and Kwon-Chung, 1998; Liu et al.,



(legend on next page)

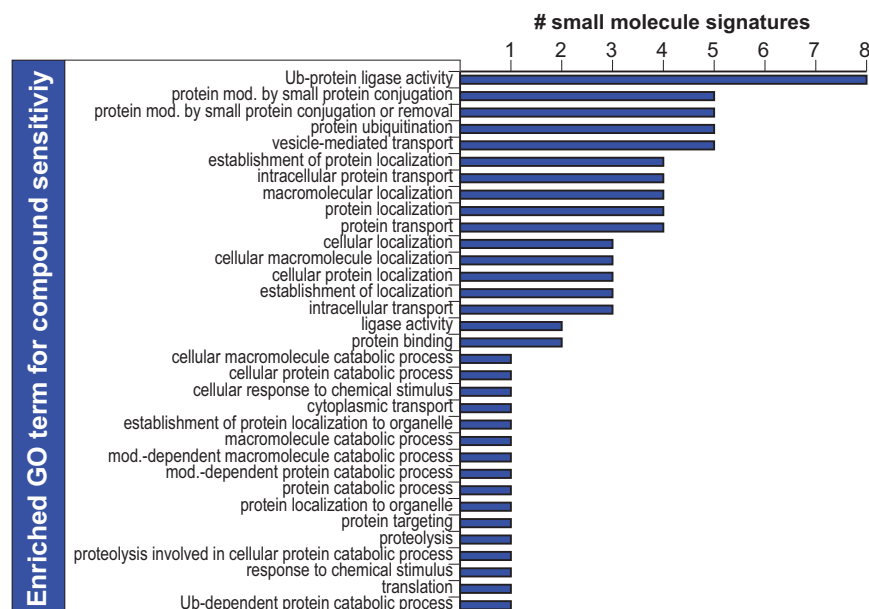


Figure 2. Determinants of Compound Sensitivity

We calculated whether molecules elicited a significant response from *C. neoformans* ORFs that are enriched for association with specific GO terms. Terms are listed on the y axis and the number of compounds whose responding gene knockouts associated with that GO term are listed on the x axis. See also Table S6.

sonication test is insufficient to definitively measure capsule maintenance. We therefore analyzed how much glucuronoxylomannan (GXM), the major capsular polysaccharide (Doering, 2009), is secreted into the growth medium by blotting with α -GXM antibodies (Figure S2A). We found that two mutants that produce little (*gcn5* Δ) or no (*yap1* Δ) visible capsule still shed GXM into the medium, suggesting that they cannot retain capsule on their cell surface.

2008). We found that seven of nine mutants in the *pbx1* Δ /*cpl1* Δ cluster exhibit a statistically significant capsule defect, as did four of the seven mutants in the *cap60* Δ cluster. In contrast, previous work from our laboratory found that approximately 1% of the original *C. neoformans* library shows a gross defect in capsule production (Liu et al., 2008).

Previous work showed that *pbx1* Δ mutants produce polysaccharide capsule whose attachment to the cell wall is sensitive to sonication, a finding that we confirmed (Figures 4C and 4D). We refer to the cell's ability to retain GXM on the cell surface as "capsule maintenance." Knockout mutants in *cnag_01058* do not exhibit a basal capsule defect but lost nearly 40% of their capsule diameter following sonication. Cells deleted for the *GCN5* gene, like *pbx1* Δ cells, show both decreased capsule levels and sonication-sensitive capsule. None of the mutants from the *cap60* Δ cluster produces a sonication-sensitive phenotype, suggesting that the *pbx1* Δ /*cpl1* Δ and *cap60* Δ clusters organize mutants that have distinct phenotypes. However, because several mutants do not produce visible capsule, the

Indeed, we found that they shed more GXM than *pbx1* Δ cells. Four of nine mutants in the *pbx1* Δ /*cpl1* Δ cluster exhibit a maintenance defect, whereas none of the *cap60* Δ cluster mutants do. We also found that GXM produced by these cells can be taken up and added to the surface ("donated") of an acapsular mutant using a standard GXM transfer assay (Kozel and Hermerath, 1984; Reese and Doering, 2003). Moreover, apparent capsule-defective mutants shed GXM (Figures S2B and S2C) and can donate GXM from conditioned medium (Figure S2C). Mutants that appear to not secrete GXM (*pbx1* Δ , *cpl1* Δ , and *sgf73* Δ) can donate it, but only if conditioned medium concentration is increased 10-fold (Figure S2D). These data are consistent with a recently published study on the role of Pbx1 in capsule attachment and assembly (Kumar et al., 2014).

Since the capsule is a major virulence trait of *C. neoformans*, we tested whether knockout mutants that exhibited a capsule defect displayed a defect in the mammalian host, using a murine inhalation model. We infected mice with a mixture of differentially-tagged wild-type and mutant cells at a ratio of 1:1. At 10 days

Figure 1. Chemical-Genetic Profiling of *C. neoformans*

- (A) Heat map of full data set following hierarchical clustering. Compounds are arrayed on the x axis and gene knockouts on the y axis. See also Tables S1 and S2.
- (B) Probability density function for pairwise correlation scores between the chemical genetic profiles of different compounds (gray) and the same compounds at different concentrations (purple) screened on different days (different batches). Scores between the chemical-genetic profiles of different concentrations of the same compounds are significantly higher than those between different compounds (Wilcoxon test, $p = 2.7 \times 10^{-176}$). See also Figure S1.
- (C) Probability density function for pairwise correlation scores between the chemical genetic profiles of different compounds (gray) and azole family compounds (purple). Pairwise comparisons between azoles exhibit higher correlation scores than nonazole compounds (Wilcoxon test, $p = 2.8 \times 10^{-6}$). Molecules with the highest pairwise comparisons scores are listed on the right.
- (D) Pearson's correlation score between two different concentrations of the same compounds. Concentrations with similar correlation scores are binned together (y axis). For compounds with the greatest correlation scores between concentrations, Venn diagrams of significant genes ($Z < -2.5$) present in profiles from the same compounds at different concentrations and the small-molecule structure are shown. The orange line indicates a hypergeometric p value ≤ 0.05 .
- (E) Histogram showing the number of deletion mutants that have given number of phenotypes. A phenotype is considered $|Z| > 2.5$ and we identified phenotypes independently for each small-molecule concentration.
- (F) Histogram showing the number of small molecules that have a given number of phenotypes. Phenotypes ($|Z| > 2.5$) were identified for each small-molecule condition/concentration.

See also Figure S1 and Tables S1, S2, S3, S4, and S5.

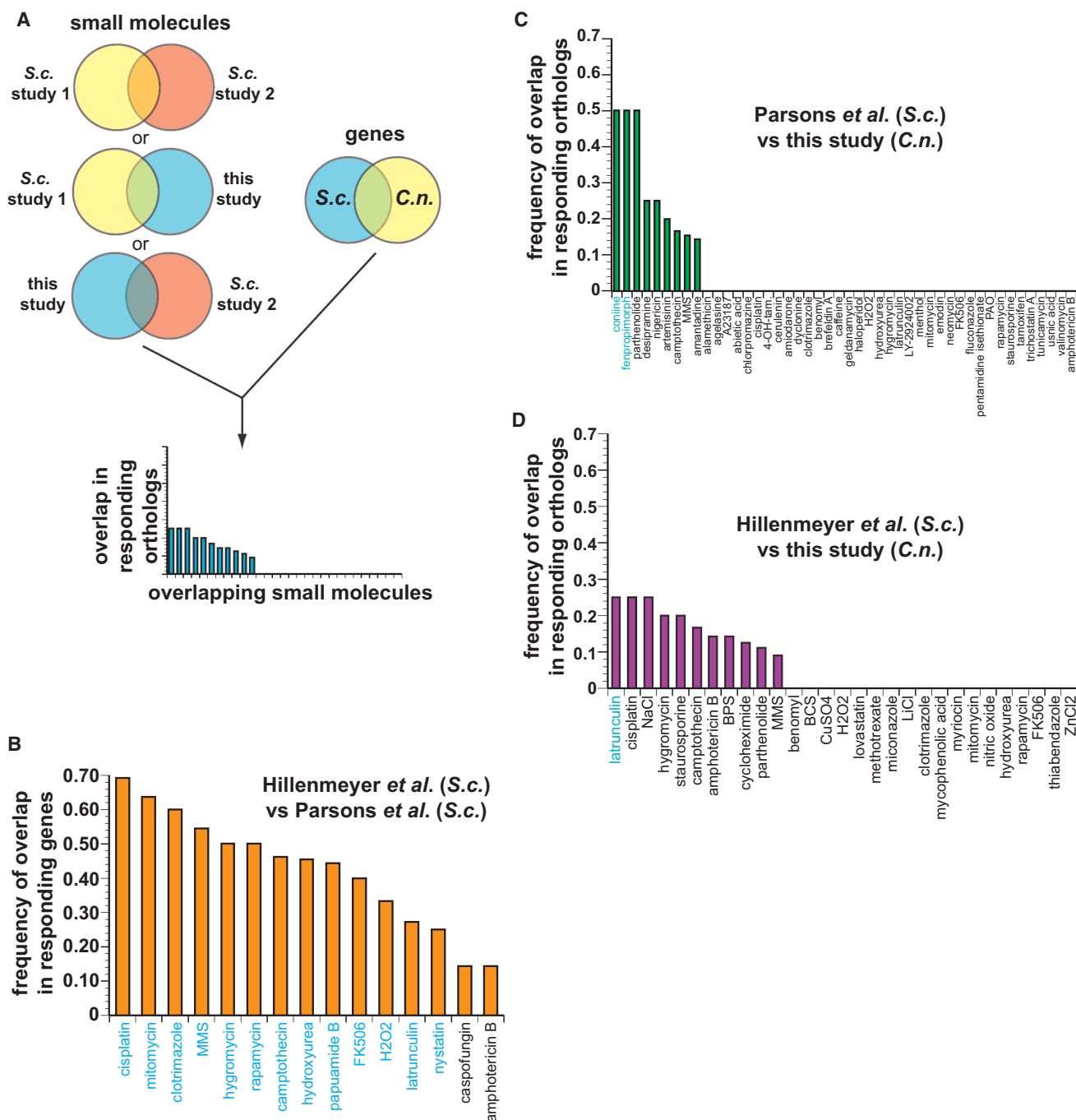


Figure 3. Chemical-Genetic Signatures of *C. neoformans* Genes Differ from Orthologous *S. cerevisiae* Genes

(A) Flowchart of computation process for comparing data sets. We identified *C. neoformans* and *S. cerevisiae* orthologous genes that were present in all data sets, then compared the responses of only those genes in all the data sets. We compared genes whose knockout mutants significantly ($|Z| > 2.5$) responded to compound that were common in at least two of the data sets.

(B) Comparison between Parsons et al. (2006) and Hillenmeyer et al. (2008), comparing the response ($|Z| > 2.5$) of genes that have orthologs present in the *C. neoformans* data set. Compounds whose profiles exhibit significant overlaps ($p < 0.05$) are labeled in blue.

(C) Comparison between our data set and Parsons et al. (2006) Compounds whose profiles exhibit significant overlaps ($p < 0.05$) are labeled in blue.

(D) Comparison between our data set and Hillenmeyer et al. (2008) Compounds whose profiles exhibit significant overlaps ($p < 0.05$) are labeled in blue.

postinfection (dpi), we sacrificed animals, harvested and homogenized lung tissue, then plated on the appropriate selective media for colony forming units (CFUs). All but one of the *pbx1Δ/cpl1Δ* cluster members were significantly underrepresented relative to wild-type; the exception was the *cnag_01058Δ* mutant, which is defective in capsule maintenance but not capsule biosynthesis (Figures 4C and S2A). *yap1Δ* cells, which appear acapsular but secrete GXM, displayed a major defect in fitness in the host (Figure 4E). Three of four *cap60Δ* cluster mutants also display a defect in accumulation of CFUs in host lungs (Figure 4E).

Chemogenomics Identifies the Cell Cycle as a Target of the Antifungal Small Molecule S8

We included a number of drug-like antifungal compounds in our screen in order to identify their targets (Table 1). Our use of *C. neoformans* chemogenomics to assist in the identification of a target of toremifene is described elsewhere (Butts et al., 2014). Here we investigate the thiazolidine-2,4-dione derivatives originally described for their activity against *C. albicans* biofilms (Kagan et al., 2014).

Our chemogenomic profiling data of the thiazolidine-2,4-dione derivative S8 revealed a striking outlier: a knockout mutant in the gene coding for a *C. neoformans* ortholog of the conserved cell-cycle kinase Wee1, is relatively resistant (Figure 5A). We observed resistance at multiple concentrations of S8 (Table S2). The related compound NA8, which contains a replacement of a sulfur atom with a carbon atom on the thiazolidinedione moiety (Figure 5B), does not elicit the same resistance (Figure S3A). The *wee1Δ* mutant is also resistant to S10 (Figure S3B), which harbors a C10 alkyl chain instead of C8 but is otherwise identical to S8 (Figure S3C).

Wee1 regulates the G2/M cell-cycle checkpoint through inhibitory phosphorylation of Cdk1, which in turn is required for cells to traverse the checkpoint. The essential phosphatase Cdc25 activates Cdk1 by removing the inhibitory phosphorylation added by Wee1 (Morgan, 2007) (Figure 5C). Because the *wee1Δ* is relatively resistant to S8, we hypothesized that S8 targeted a protein that acts through Wee1 to regulate Cdk1. One such target could be Cdc25.

We reasoned that if the Wee1/Cdc25-regulated step of the cell cycle were an important target of S8 in vivo, wild-type *C. neoformans* cells treated with S8 would arrest at G2/M. To test this prediction, we treated exponential cultures with S8, S10, or NA8 and examined the impact on the cell cycle. We harvested and fixed representative samples every 30 min, then analyzed DNA content by flow cytometry. Control cultures treated with DMSO (carrier) (Figure 5D) or the control compound NA8 (Figure 5E) stayed asynchronous for the entire 3.5 hr of the time course. Strikingly, S8-treated (Figure 5F) cells accumulated with 2C DNA content, which indicates G2/M arrest in *C. neoformans*, a haploid yeast (Whelan and Kwon-Chung, 1986). At later time points, cells synthesize DNA but do not complete mitosis and cytokinesis. This is consistent with observations in *S. pombe* that partial inhibition of Cdk1 permits replication of DNA (Broek et al., 1991).

Because inhibition of Cdc25 would provide a parsimonious explanation for the genetic and biological properties of S8, we tested whether S8 inhibits *C. neoformans* Cdc25 in vitro. We ex-

pressed and purified the catalytic domain of a *C. neoformans* ortholog (CNAG_07942) in *E. coli* (Figure S3D) and then performed in vitro phosphatase assays using 3-O-methyl fluorescein phosphate (OMFP) as a substrate (Figures 5G and 5H) (Hill et al., 1968). We observed that S8 inhibits Cdc25 activity ($K_i \sim 140 \mu\text{M}$, Figure 5E), as do both S10 (Figure S3E) and NSC 663284 ($K_i \sim 250 \mu\text{M}$, Figure S3F), a commercially available inhibitor of mammalian Cdc25 (Pu et al., 2002). The control compound NA8 does not inhibit *C. neoformans* Cdc25 in vitro (Figure S3G). For S8, the in vitro inhibition constant is roughly comparable to the liquid MIC value against *C. neoformans*, which we measured to be $\sim 60 \mu\text{M}$ in YNB. S10 has a higher K_i ($K_i \sim 310 \mu\text{M}$) but similar to the MIC value ($\sim 55 \mu\text{M}$) measured in YNB agar compared to S8.

O2M: A Genetic Biomarker Algorithm to Predict Compound Synergies

Drug resistance is a major clinical challenge in the treatment of both bacterial and fungal infections (Anderson, 2005; Cantas et al., 2013). An effective therapeutic strategy is to treat patients with drugs that act synergistically, enhancing each other's effectiveness beyond that produced by the sum of each drug's individual impact (Kalan and Wright, 2011). This approach is thought to decrease acquisition drug resistance, increase the available drug repertoire (Kalan and Wright, 2011) and ameliorate toxicities (Kathiravan et al., 2012; Lehár et al., 2009).

We hypothesized that we could use the chemogenomic information from our screens of drugs known to act synergistically, such as FLC and fenpropimorph (Jansen et al., 2009), to identify new synergistic interactions (Figure 6A). When we compared the identity of genes whose knockouts "responded" to each individual small molecule in a known synergistic pair ($|Z| \geq 2.5$, Tables S3 and S4), we found that this "responsive" gene set was significantly enriched over the expected value (Fisher's exact test, $p \leq 6 \times 10^{-5}$) (Figure 6A, top). This observation is consistent with a previous report that the chemical-genetic response to each drug in a synergistic pair is enriched for overlapping genes (Jansen et al., 2009).

This overlap in responsive gene sets led us to consider the possibility that overlapping responsive genes from known synergistic compound pairs could be used as biomarkers to predict new synergistic combinations. Our method involves first identifying the overlaps in responsive gene sets for all compounds that had been reported in the literature to synergize with a small molecule of interest ("compound X"), selecting those genes common to all of those sets (Figure 6A, middle, the overlaps of overlaps). We refer to these genes as "synergy biomarker genes." Critically, we next hypothesized that any compound that contains one or more of these synergy biomarker genes in its responsive gene set would be synergistic with compound X. Because our method used the overlaps of response gene overlaps between compounds known to be synergistic, we refer to it as the "overlap-squared method" or "O2M."

We then tested O2M using two drugs for which substantial literature synergy information was available: FLC and geldanamycin (GdA). FLC is an approved antifungal drug. GdA is an inhibitor of Hsp90, a chaperone protein with many physical and genetic interactions (Taipale et al., 2010). We performed our analysis on fenpropimorph and sertraline, which are known to

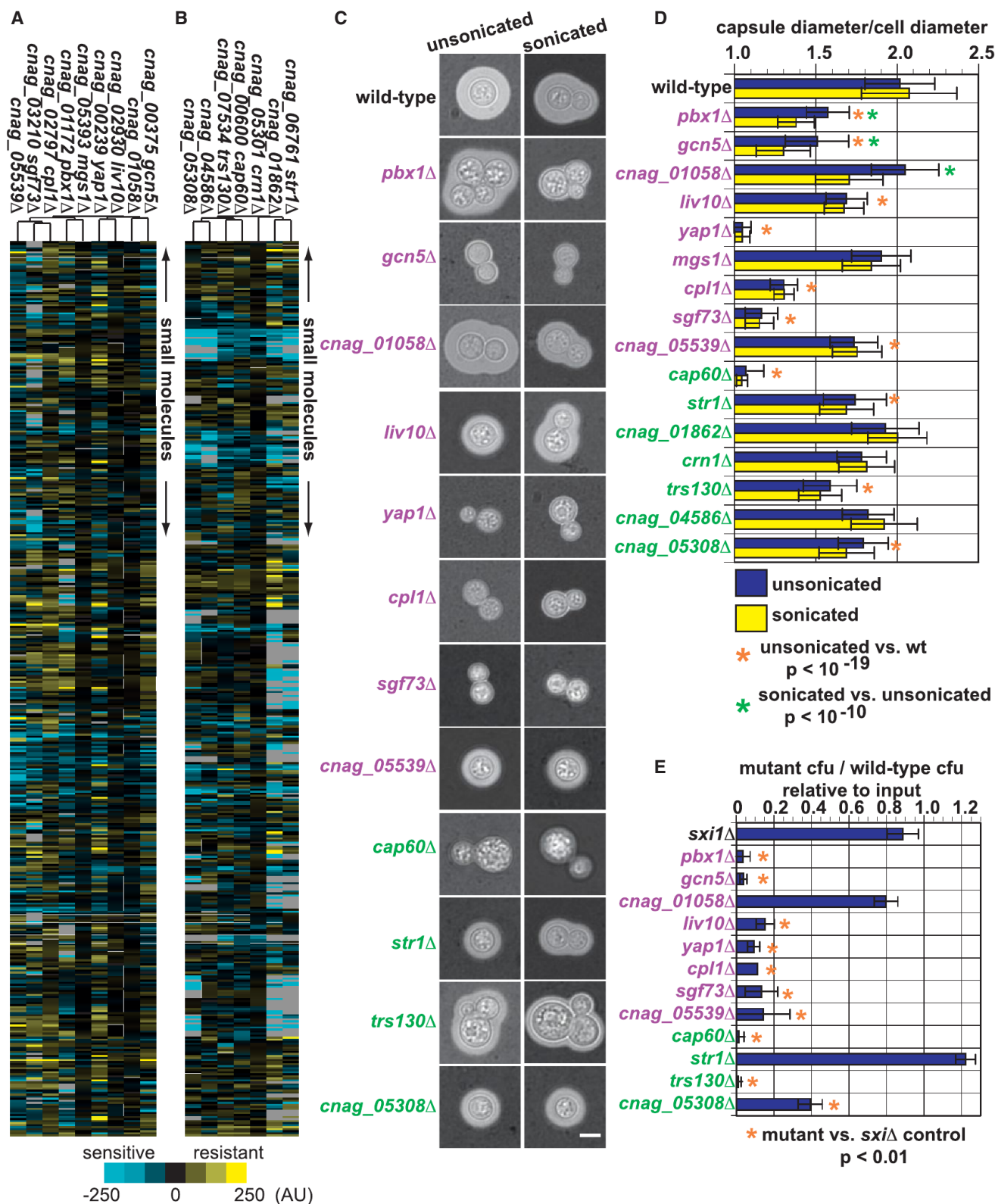


Figure 4. Chemical-Genetic Profiling Identifies Genes Involved in Capsule Biosynthesis

(A) Cluster containing the chemical signatures of the *pbx1*Δ and *cpl1*Δ mutants.
(B) Cluster containing the chemical signatures of the *cap60*Δ mutants.

(legend continued on next page)

act synergistically with FLC (Jansen et al., 2009; Zhai et al., 2012), and cyclosporine and rapamycin, which are known to act synergistically with GdA (Francis et al., 2006; Kumar et al., 2005). Using this prior knowledge and our data, we identified synergy biomarker genes for FLC (CNAG_00573, CNAG_03664, and CNAG_03917) and GdA (CNAG_01172, CNAG_03829, and CNAG_01862). We generated a list of compounds from our chemical-genetics data set that contain one or more of these genes in their responsive genes set.

We then used a standard “checkerboard” assay to experimentally determine fractional inhibitory concentration index (FICI), and we adopted the standard that an FICI value below 0.5 is synergistic (Meletiadiis et al., 2010). We determined FICIs for FLC and GdA with three sets of compounds: (1) the compounds predicted from synergy biomarker genes, (2) the predicted synergistic compounds for the other drug (e.g., we tested compounds predicted to be synergistic with GdA for synergy with FLC), and (3) a randomly generated subset of the compounds not predicted to act synergistically with either FLC or GdA. The second and third groups are as controls for compounds that are generally synergistic and to determine the background frequency of synergistic interactions within a set of compounds.

Respective experimental FICI values for FLC and GdA are shown in Figures 6B and 6C (yellow bars: synergy; blue bars additive or worse interactions). The labels for compounds we predicted to be synergistic are colored purple, positive controls (published synergistic compound pairs) are colored green, and predicted negative control compounds are colored blue (Figure 6). We observed that only ~10% of the negative control compounds act synergistically with either FLC or GdA. In striking contrast, we found ~80% and ~60% of the compounds selected by O2M are synergistic with FLC and GdA, respectively. Thus, for two unrelated compounds, O2M is highly successful at predicting synergistic interactions and performs vastly better than the brute force trial-and-error approach (Figures 6D and 6E) ($p < 0.0008$, Fisher’s exact test).

DISCUSSION

We applied chemogenomic profiling to the major fungal driver of AIDS-related death, the encapsulated yeast *C. neoformans*, to produce a chemical-genetic atlas of this important pathogen. Beyond identifying new virulence factors and compound mode of action, we describe a conceptually general approach to identifying drug synergies that combines prior knowledge and chemogenomic profiles.

A Chemical-Genetic Atlas for *C. neoformans*

We maximized the quality of the atlas in several ways. To capture concentration-dependent impacts of compounds, we obtained the MIC for each compound and examined the genetic re-

sponses at multiple concentrations below MIC. In addition, we performed a large number of control screens and incorporated batch information for systematic correction. Overall benchmarks of data quality (Figure 1) together with nearest neighbor and Gene Ontology analysis (Figure 2) support the existence of substantial chemical-genetic signal in the data. Even genes with orthologs in both *S. cerevisiae* and *C. neoformans* show considerable differences in responses (Figure 3). While this may not be surprising given the large phylogenetic distance between these fungi, it shows that understanding the chemical responses of pathogens requires pathogen-focused studies, even when considering conserved genes and processes. For example, we observed differences in the responses to azole drugs between *S. cerevisiae* and *C. neoformans* (Figure 3). Since azoles are heavily used clinically, differences in responses between species are of significant interest.

Insights Gained from Initial Use of the *C. neoformans* Chemical-Genetic Atlas

Identification of Mutants that Impact Capsule Formation and Mammalian Infection

Our studies on capsule biosynthesis genes focused two different clusters that contained genes that we and others have shown to be required for capsule formation, the *pbx1/cpl1Δ* cluster and the *cap60Δ* cluster. As anticipated from model organism studies (Collins et al., 2007; Costanzo et al., 2010; Nichols et al., 2011; Parsons et al., 2004; Parsons et al., 2006), these clusters were indeed enriched for genes whose mutants are defective in capsule biosynthesis and mammalian pathogenesis. The genes represented by the two clusters differed functionally in that genes in the *pbx1/cpl1Δ* cluster but not the *cap60Δ* cluster are required for association of capsule polysaccharide with the cell surface (Figures 4 and S2). A recent study on Pbx1 and its ortholog, Pbx2, proposes that the two proteins act redundantly in capsule assembly (Kumar et al., 2014). *pbx1Δ* and *pbx2Δ* cells shed lower amounts of GXM into the culture medium but that the GXM functions in a capsule transfer assay. Electron microscopy studies indicate that these mutants exhibit defects in the cell wall. Our data are fully consistent with these data. Other genes from the *pbx1Δ/cpl1Δ* cluster likely play a role in these processes. Some, like *GCN5* and *SGF73*, which encode orthologs of the yeast SAGA histone acetylase/deubiquitylase complex, are clearly regulatory, while others could act more directly. While detailed validation and investigation of these many candidates (including gene deletion reconstruction studies) will be required to obtain mechanistic insight into capsule biology, their enrichment suggests value of this Cryptococcal chemogenomic resource in identifying mutants defective in virulence.

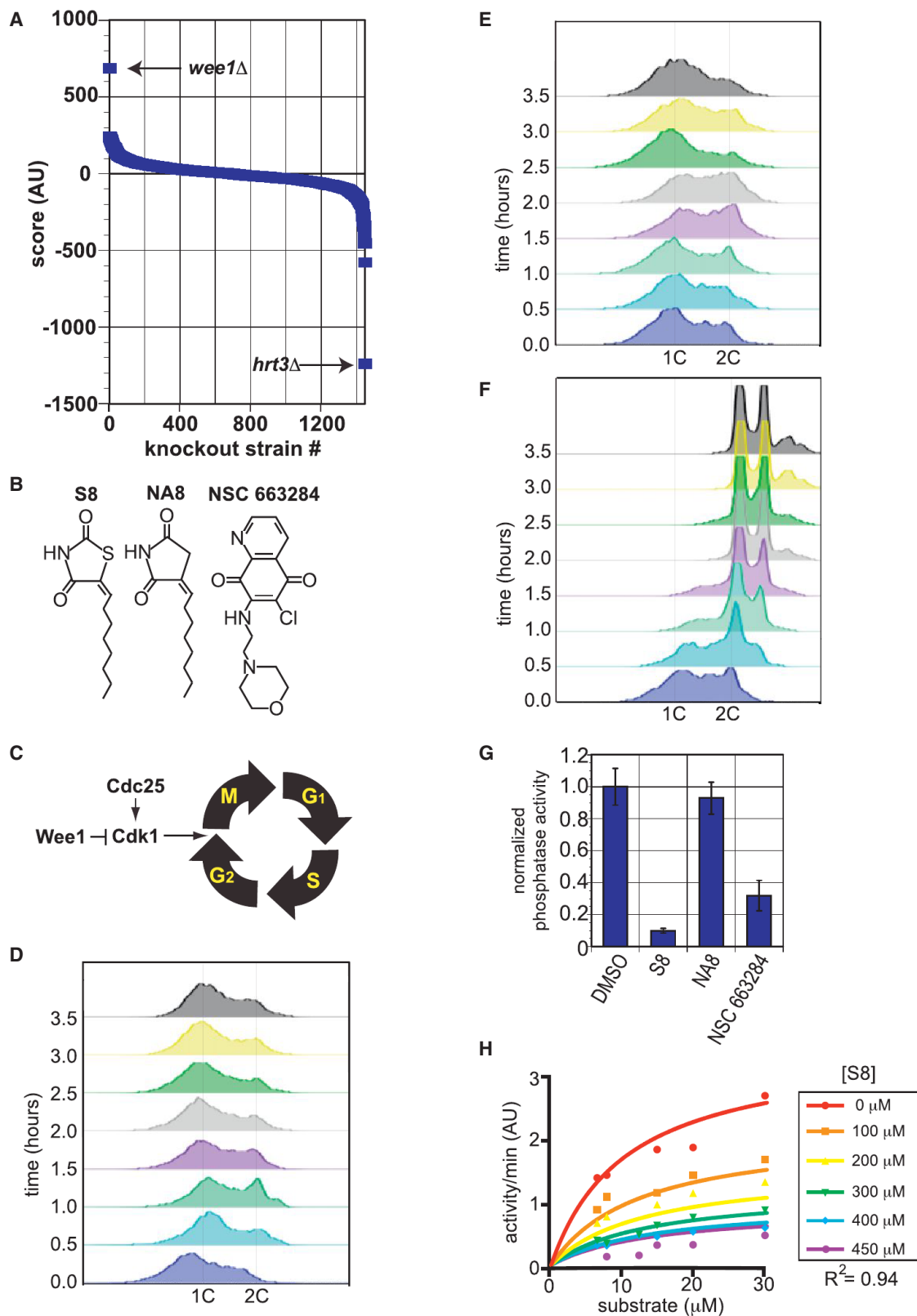
Compound Target Identification

Chemogenomic profiling has proven useful in identifying targets of uncharacterized compounds (Parsons et al., 2006), including

(C) Images of individual cells grown in 10% Sabouraud’s broth to induce capsule. Representative cells are shown for mutants that exhibit a statistically significant phenotype. Scale bar, 5 μ m.

(D) Quantification of capsule sizes from all mutants in *pbx1Δ/cpl1Δ* (purple labels) cluster or *cap60Δ* (green labels) cluster. 100 cells were measured for each strain, the error bars represents the standard deviation, and p values were calculated using Student’s t test.

(E) Colony counts from colony forming units (cfu) extracted from mouse lungs following an inhalation infection. Three mice are shown for each datapoint; the error bars represent the standard deviation and p values were calculated using Student’s t test.



(legend on next page)

in the pathogenic fungus *C. albicans* (Jiang et al., 2008; Xu et al., 2007; Xu et al., 2009). Chemical-genetic data can be used to determine the target of compounds within complex mixtures (Jiang et al., 2008; Xu et al., 2009). Our goal differed: we sought to identify targets of repurposed compounds, as described elsewhere (Butts et al., 2013), or, in the case of S8, a compound identified as an inhibitor of *Candida* biofilms (Kagan et al., 2014). The identification of the Wee1 kinase as a sensitivity determinant for S8, the cell-cycle arrest produced by S8, and the ability of the compound to inhibit *Cn*Cdc25 in vitro together support the model that S8 inhibits growth through via the cell cycle at least in part via inhibition of Cdc25. Whether this explains its impact on biofilms requires further investigation. As with any compound target, ultimate proof that Cdc25 is the target of S8 will require the isolation of resistance alleles of *CDC25*.

Given the simplicity of the pharmacophore and its K_i for *Cn*Cdc25, it would not be surprising if S8 had additional cellular targets, as recently described (Feldman et al., 2014). Cdc25 is a conserved cell-cycle phosphatase and therefore might be considered a poor drug target a priori but cyclin-dependent kinases are a focus of recent antiparasite therapeutics (Geyer et al., 2005). It is also notable that the target of azole antifungals, lanosterol 14-demethylase (Ghannoum and Rice, 1999) is conserved from yeast to human.

O2M: Predicting Compound Synergies Using Prior Knowledge and Chemical Profiles

Identifying synergistic drug interactions is of considerable clinical interest, but efficient methods for their identification are elusive. Systematic examination of combinations of a small set of compounds using *S. cerevisiae* suggests that synergies are relatively rare and often involve so-called “promiscuous” synergizers, compounds that are synergistic with multiple partners (Cokol et al., 2011). Chemogenomic studies have shown that drugs known to be synergistic tend to have overlapping “responding” gene sets (Jansen et al., 2009). We expanded on this concept to develop a highly parallel method, O2M, for efficiently predicting synergistic drug interactions. Our work utilizes prior knowledge of drug synergies to identify a discrete set of predictive biomarker genes for a given compound. We experimentally demonstrated the utility of O2M for two compounds, FLC and geldanamycin. Our method identified dozens of synergistic interactions and discovered a small number of biomarkers that could serve as readouts for further screens for synergistic drugs. The method appears to not simply select promiscuous synergizers: five of six drugs previously classified as promiscu-

ous synergizers (Cokol et al., 2011) were tested in our studies but most were not predicted to be synergistic by O2M. One of the promiscuous compounds was a positive control (fenpropimorph with FLC) and another (dyclonine) was predicted synergistic with FLC but was not and was predicted not synergistic with GdA but was. We anticipate that O2M could be used to identify synergistic compound interactions in published *E. coli* and *C. albicans* chemical-genetics data sets (Jiang et al., 2008; Nichols et al., 2011; Xu et al., 2007; Xu et al., 2009).

EXPERIMENTAL PROCEDURES

Determination of MICs

We determined MIC on solid growth medium for each compound used in screening (Table 1).

Colony Array-Based Chemogenomic Profiling

C. neoformans knockouts were inoculated from frozen 384-well plates to YNB + 2% glucose. Plates were grown 24 hr at 30°C, then used to inoculate screening plates containing compounds of interest.

Data Analysis

Data were analyzed as previously described (Baryshnikova et al., 2010) with the a few exceptions.

***C. neoformans* Ortholog Identification and GO Term Mapping**

Mapping from *S. cerevisiae* Uniprot Proteins to *C. neoformans* Uniprot Proteins was done using One-to-one mappings in MetaPhOrs (<http://metaphors.phylomedb.org/>). *C. neoformans* ORFs were compared to a database of *S. cerevisiae* Uniprot Proteins using blastp (Altschul et al., 1997) with a E-score cutoff of 10^{-30} . Corresponding yeast GO annotations were mapped onto the *C. neoformans* ORFs.

Comparison of Transcriptional Response to FLC

Compared transcriptional responses between *S. cerevisiae* (Kuo et al., 2010) and *C. neoformans* (Florio et al., 2011).

Capsule Induction Assay

Samples were grown overnight at 30°C in 100% Sabouraud's broth, then diluted 1:100 into 10% Sabouraud's broth buffered with 50 mM HEPES pH 7.3 and grown for 3 days at 37°C. India ink was added at 3:1 ratio and samples imaged on a Zeiss Axiovert microscope.

Capsule Transfer Assay

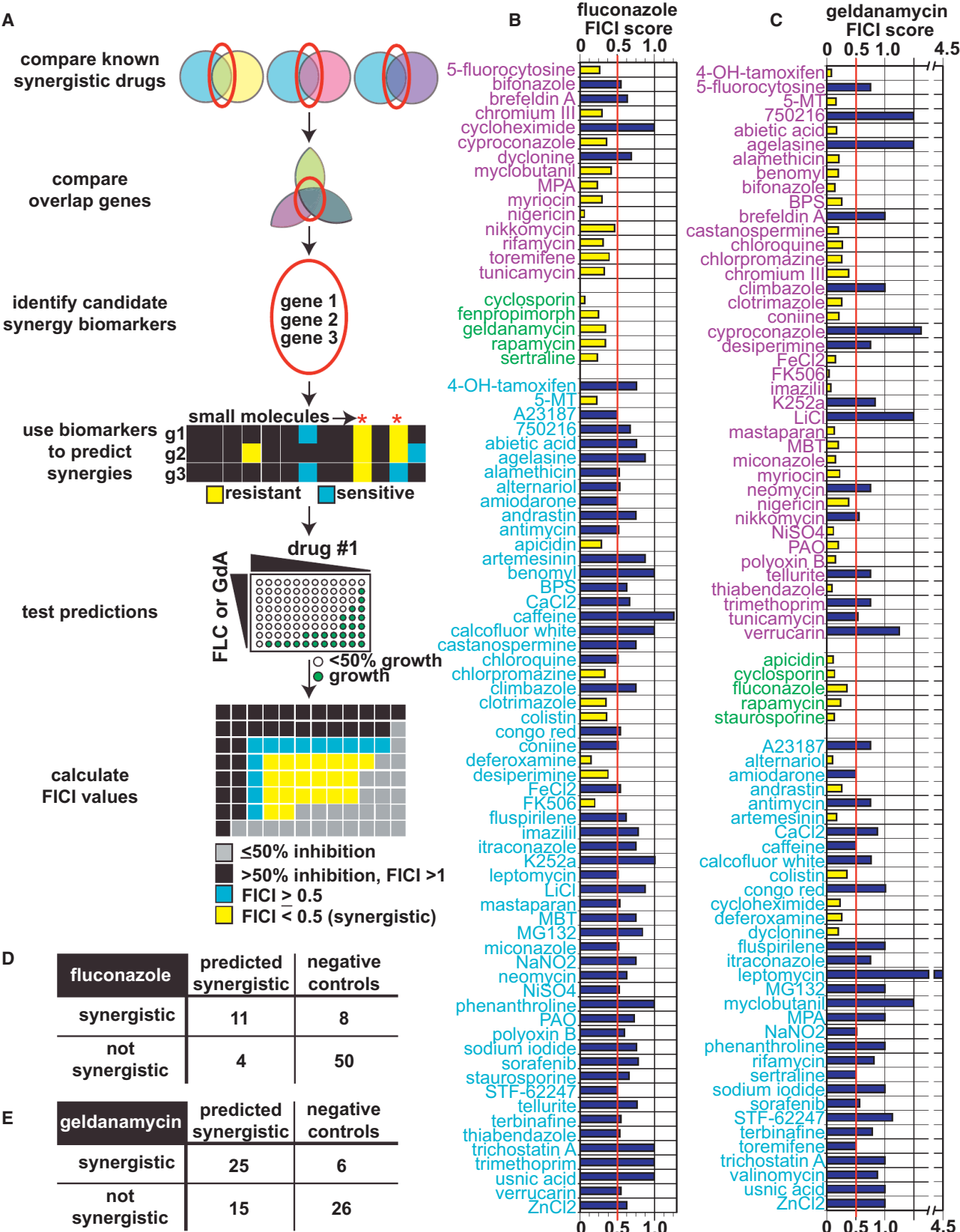
Performed as in (Reese and Doering, 2003), with minor modifications.

GXM Immunoblot Assay

Conditioned medium was made from donor GXM donor strains as described above.

Figure 5. *C. neoformans* Cdc25 Is a Target of S8 In Vivo and In Vitro

- Chemical-genetic data of the growth scores of each knockout mutant grown on S8 (y axis). The mutant that exhibited the greatest resistance is *wee1Δ*. The mutant strain that showed the greatest sensitivity to S8 is *cnag_04462Δ*.
- Structures of S8, NA8, and NSC 663284. The structure of S10 is shown in Figure S3C.
- G2/M regulation (Morgan, 2007).
- DNA content of asynchronous *C. neoformans* culture split into aliquots for treatment with compounds of interest, with samples harvested at appropriate times. Data for DMSO-treated culture is shown.
- DNA content from NA8-treated culture from same starting culture as Figure 5F.
- DNA content from S8-treated culture from same starting culture as Figure 5F.
- Phosphatase activity of purified *C. neoformans* Cdc25 catalytic domain (CNAG_01572, aa442-662). Average of three independent replicates are shown and the error bars represent the standard deviation.
- Michaelis-Menten kinetics of S8 inhibition of *Cn*Cdc25 from in vitro phosphatase activity. A noncompetitive model of enzyme inhibition produced the best R^2 value (0.94).



(legend on next page)

Mouse Infection Assay

Mouse lung infections were performed as previously described (Chun et al., 2011).

Cdc25 Protein Purification

We identified the *C. neoformans* ortholog of Cdc25, *CNAG_01572*, by best reciprocal BLAST (Altschul et al., 1997) hit with the human Cdc25A, Cdc25B, and Cdc25C protein isoforms. We then inserted the exonic sequence of the catalytic domain into a 6x-His tag expression vector for purification.

Cdc25 Phosphatase Assay

Cdc25 phosphatase activity was analyzed in activity buffer (50 mM Tris pH 8.3, 5% glycerol, 0.8 mM dithiothreitol, and 1% PVA).

Cdc25 Inhibitor Treatment and FACS Analysis

Wild-type *C. neoformans* was grown overnight in 1x YNB at 30°C with rotation. Cultures were diluted to OD₆₀₀ ~0.2 into 150 ml 1x YNB, then incubated 3 hr at 30°C. Samples were then split and NA8, S8, and S10 added to 60 μM. Equivalent volume of DMSO was added to the control culture.

Fractional Inhibitory Concentration Index Assay for Synergy

We determined FICI using a standard checkerboard assay (Hsieh et al., 1993), calculating FICI as described using a 50% growth inhibition cutoff for MICs for individual compounds (Hsieh et al., 1993; Meletiadi et al., 2010), then using a standard cutoff of FICI < 0.5 to define synergy.

See Extended Experimental Procedures for additional details.

SUPPLEMENTAL INFORMATION

Supplemental Information includes Extended Experimental Procedures, three figures, and seven tables and can be found with this article online at <http://dx.doi.org/10.1016/j.cell.2014.10.044>.

AUTHOR CONTRIBUTIONS

J.C.S.B. and H.D.M. designed the study. J.C.S.B. carried out all of experiments described in the paper. J.N. and C.L.M. performed data analysis shown in Figures 1B–1D, Figure 2 and Figure 3. B.V., R.D., and C.L.M. filtered, de-noised and scored the primary colony array data. A.B., S.K., I.P., and D.J.K. provided compounds and guidance. J.C.S.B. and H.D.M. wrote the manuscript with input from all coauthors.

ACKNOWLEDGMENTS

This work was supported by NIH grants 5R01AI099206 to H.D.M., 1R01AI091422 to D.J.K., 1R01HG005084 and 1R01GM04975 to C.L.M., and a grant from the CIFAR Genetic Networks Program to C.L.M.

Received: July 3, 2014

Revised: August 28, 2014

Accepted: October 22, 2014

Published: November 20, 2014

REFERENCES

- Altschul, S.F., Madden, T.L., Schäffer, A.A., Zhang, J., Zhang, Z., Miller, W., and Lipman, D.J. (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* 25, 3389–3402.
- Anderson, J.B. (2005). Evolution of antifungal-drug resistance: mechanisms and pathogen fitness. *Nat. Rev. Microbiol.* 3, 547–556.
- Armstrong-James, D., Meintjes, G., and Brown, G.D. (2014). A neglected epidemic: fungal infections in HIV/AIDS. *Trends Microbiol.* 22, 120–127.
- Baryshnikova, A., Costanzo, M., Kim, Y., Ding, H., Koh, J., Toufighi, K., Youn, J.-Y., Ou, J., San Luis, B.-J., Bandyopadhyay, S., et al. (2010). Quantitative analysis of fitness and genetic interactions in yeast on a genome scale. *Nat. Methods* 7, 1017–1024.
- Broek, D., Bartlett, R., Crawford, K., and Nurse, P. (1991). Involvement of p34cdc2 in establishing the dependency of S phase on mitosis. *Nature* 349, 388–393.
- Butts, A., and Krysan, D.J. (2012). Antifungal drug discovery: something old and something new. *PLoS Pathog.* 8, e1002870.
- Butts, A., DiDone, L., Koselny, K., Baxter, B.K., Chabrier-Rosello, Y., Wellington, M., and Krysan, D.J. (2013). A repurposing approach identifies off-patent drugs with fungicidal cryptococcal activity, a common structural chemotype, and pharmacological properties relevant to the treatment of cryptococcosis. *Eukaryot. Cell* 12, 278–287.
- Butts, A., Koselny, K., Chabrier-Roselló, Y., Semighini, C.P., Brown, J.C.S., Wang, X., Annadurai, S., DiDone, L., Tabroff, J., Childers, W.E., Jr., et al. (2014). Estrogen receptor antagonists are anti-cryptococcal agents that directly bind EF hand proteins and synergize with fluconazole in vivo. *MBio* 5, e00765–e13.
- Cantas, L., Shah, S.Q.A., Cavaco, L.M., Manaia, C.M., Walsh, F., Popowska, M., Garelick, H., Bürgmann, H., and Sørum, H. (2013). A brief multi-disciplinary review on antimicrobial resistance in medicine and its linkage to the global environmental microbiota. *Front Microbiol.* 4, 96.
- Chang, Y.C., and Kwon-Chung, K.J. (1998). Isolation of the third capsule-associated gene, CAP60, required for virulence in *Cryptococcus neoformans*. *Infect. Immun.* 66, 2230–2236.
- Chun, C.D., Brown, J.C.S., and Madhani, H.D. (2011). A major role for capsule-independent phagocytosis-inhibitory mechanisms in mammalian infection by *Cryptococcus neoformans*. *Cell Host Microbe* 9, 243–251.
- Cokol, M., Chua, H.N., Tasan, M., Mutlu, B., Weinstein, Z.B., Suzuki, Y., Nergiz, M.E., Costanzo, M., Baryshnikova, A., Giaever, G., et al. (2011). Systematic exploration of synergistic drug pairs. *Mol. Syst. Biol.* 7, 544.
- Collins, S.R., Miller, K.M., Maas, N.L., Roguev, A., Fillingham, J., Chu, C.S., Schuldiner, M., Gebbia, M., Recht, J., Shales, M., et al. (2007). Functional dissection of protein complexes involved in yeast chromosome biology using a genetic interaction map. *Nature* 446, 806–810.
- Costanzo, M., Baryshnikova, A., Bellay, J., Kim, Y., Spear, E.D., Sevier, C.S., Ding, H., Koh, J.L.Y., Toufighi, K., Mostafavi, S., et al. (2010). The genetic landscape of a cell. *Science* 327, 425–431.
- Del Poeta, M. (2004). Role of phagocytosis in the virulence of *Cryptococcus neoformans*. *Eukaryot. Cell* 3, 1067–1075.

Figure 6. O2M Approach for Predicting Compound Synergy

(A) Approach for predicting compound synergistic interaction.

(B) FICI values for fluconazole (FLC). Predicted synergistic compounds are labeled in purple and known synergistic compounds in green. Bars represent the average of two assays but both had to be FICI < 0.5 to be considered synergistic. Compounds labeled in blue are negative controls from one of two categories: 1) predicted to synergize with geldanamycin (GdA) but not FLC or 2) randomly generated list of compounds not predicted to be synergistic with either FLC or GdA. Yellow bars represent an FICI < 0.5 (synergistic) and blue bars and FICI ≥ 0.5 (not synergistic).

(C) FICI values for GdA. Labels and colors are analogous to those in part B.

(D) Contingency table of synergistic versus nonsynergistic interactions with FLC. $p < 0.0008$ (Fisher's exact test).

(E) Contingency table of synergistic versus nonsynergistic interactions with GdA. $p < 0.0008$ (Fisher's exact test).

- Dittmar, J.C., Reid, R.J., and Rothstein, R. (2010). ScreenMill: a freely available software suite for growth measurement, analysis and visualization of high-throughput screen data. *BMC Bioinformatics* 11, 353–363.
- Doering, T.L. (2009). How sweet it is! Cell wall biogenesis and polysaccharide capsule formation in *Cryptococcus neoformans*. *Annu. Rev. Microbiol.* 63, 223–247.
- Feldman, M., Al-Quntar, A., Polacheck, I., Friedman, M., and Steinberg, D. (2014). Therapeutic potential of thiazolidinedione-8 as an antibiofilm agent against *Candida albicans*. *PLoS ONE* 9, e93225.
- Fernández, L., and Hancock, R.E.W. (2012). Adaptive and mutational resistance: role of porins and efflux pumps in drug resistance. *Clin. Microbiol. Rev.* 25, 661–681.
- Florio, A.R., Ferrari, S., De Carolis, E., Torelli, R., Fadda, G., Sanguinetti, M., Sanglard, D., and Posteraro, B. (2011). Genome-wide expression profiling of the response to short-term exposure to fluconazole in *Cryptococcus neoformans* serotype A. *BMC Microbiol.* 11, 97.
- Francis, L.K., Alsayed, Y., Leleu, X., Jia, X., Singha, U.K., Anderson, J., Timm, M., Ngo, H., Lu, G., Huston, A., et al. (2006). Combination mammalian target of rapamycin inhibitor rapamycin and HSP90 inhibitor 17-allylamino-17-demethoxygeldanamycin has synergistic activity in multiple myeloma. *Clin. Cancer Res.* 12, 6826–6835.
- Geyer, J.A., Prigge, S.T., and Waters, N.C. (2005). Targeting malaria with specific CDK inhibitors. *Biochim. Biophys. Acta* 1754, 160–170.
- Ghannoum, M.A., and Rice, L.B. (1999). Antifungal agents: mode of action, mechanisms of resistance, and correlation of these mechanisms with bacterial resistance. *Clin. Microbiol. Rev.* 12, 501–517.
- Haynes, B.C., Skowrya, M.L., Spencer, S.J., Gish, S.R., Williams, M., Held, E.P., Brent, M.R., and Doering, T.L. (2011). Toward an integrated model of capsule regulation in *Cryptococcus neoformans*. *PLoS Pathog.* 7, e1002411.
- Heitman J., Kozel T.R., Kwon-Chung K.J., Perfect J.R., and Casadevall A., eds. (2011). *Cryptococcus: from human pathogen to model yeast* (Washington, D.C.: ASM Press).
- Hill, H.D., Summer, G.K., and Waters, M.D. (1968). An automated fluorometric assay for alkaline phosphatase using 3-O-methylfluorescein phosphate. *Anal. Biochem.* 24, 9–17.
- Hillenmeyer, M.E., Fung, E., Wildenhain, J., Pierce, S.E., Hoon, S., Lee, W., Proctor, M., St Onge, R.P., Tyers, M., Koller, D., et al. (2008). The chemical genomic portrait of yeast: uncovering a phenotype for all genes. *Science* 320, 362–365.
- Hsieh, M.H., Yu, C.M., Yu, V.L., and Chow, J.W. (1993). Synergy assessed by checkerboard. A critical analysis. *Diagn. Microbiol. Infect. Dis.* 16, 343–349.
- Janbon, G., Ormerod, K.L., Paulet, D., Byrnes, E.J., 3rd, Yadav, V., Chatterjee, G., Mullapudi, N., Hon, C.-C., Billmyre, R.B., Brunel, F., et al. (2014). Analysis of the genome and transcriptome of *Cryptococcus neoformans* var. *grubii* reveals complex RNA expression and microevolution leading to virulence attenuation. *PLoS Genet.* 10, e1004261.
- Jansen, G., Lee, A.Y., Epp, E., Fredette, A., Surprenant, J., Harcus, D., Scott, M., Tan, E., Nishimura, T., Whiteway, M., et al. (2009). Chemogenomic profiling predicts antifungal synergies. *Mol. Syst. Biol.* 5, 338.
- Jiang, B., Xu, D., Allocco, J., Parish, C., Davison, J., Veillette, K., Sillaots, S., Hu, W., Rodriguez-Suarez, R., Trosok, S., et al. (2008). PAP inhibitor with in vivo efficacy identified by *Candida albicans* genetic profiling of natural products. *Chem. Biol.* 15, 363–374.
- Kagan, S., Jabbour, A., Sionov, E., Alquntar, A.A., Steinberg, D., Srebniak, M., Nir-Paz, R., Weiss, A., and Polacheck, I. (2014). Anti-*Candida albicans* biofilm effect of novel heterocyclic compounds. *J. Antimicrob. Chemother.* 69, 416–427.
- Kalan, L., and Wright, G.D. (2011). Antibiotic adjuvants: multicomponent anti-infective strategies. *Expert Rev. Mol. Med.* 13, e5.
- Kathiravan, M.K., Salake, A.B., Chothe, A.S., Dudhe, P.B., Watode, R.P., Mukta, M.S., and Gadhwane, S. (2012). The biology and chemistry of antifungal agents: a review. *Bioorg. Med. Chem.* 20, 5678–5698.
- Kozel, T.R., and Hermerath, C.A. (1984). Binding of cryptococcal polysaccharide to *Cryptococcus neoformans*. *Infect. Immun.* 43, 879–886.
- Kozel, T.R., Levitz, S.M., Dromer, F., Gates, M.A., Thorkildson, P., and Janbon, G. (2003). Antigenic and biological characteristics of mutant strains of *Cryptococcus neoformans* lacking capsular O acetylation or xylosyl side chains. *Infect. Immun.* 71, 2868–2875.
- Kumar, R., Musiyenko, A., and Barik, S. (2005). Plasmodium falciparum calcineurin and its association with heat shock protein 90: mechanisms for the anti-malarial activity of cyclosporin A and synergism with geldanamycin. *Mol. Biochem. Parasitol.* 141, 29–37.
- Kumar, P., Yang, M., Haynes, B.C., Skowrya, M.L., and Doering, T.L. (2011). Emerging themes in cryptococcal capsule synthesis. *Curr. Opin. Struct. Biol.* 21, 597–602.
- Kumar, P., Heiss, C., Santiago-Tirado, F.H., Black, I., Azadi, P., and Doering, T.L. (2014). Pbx proteins in *Cryptococcus neoformans* cell wall remodeling and capsule assembly. *Eukaryot. Cell* 13, 560–571.
- Kuo, D., Tan, K., Zinman, G., Ravasi, T., Bar-Joseph, Z., and Ideker, T. (2010). Evolutionary divergence in the fungal response to fluconazole revealed by soft clustering. *Genome Biol.* 11, R77.
- Lehár, J., Krueger, A.S., Avery, W., Heilbut, A.M., Johansen, L.M., Price, E.R., Rickles, R.J., Short, G.F., 3rd, Staunton, J.E., Jin, X., et al. (2009). Synergistic drug combinations tend to improve therapeutically relevant selectivity. *Nat. Biotechnol.* 27, 659–666.
- Liu, C., Apodaca, J., Davis, L.E., and Rao, H. (2007a). Proteasome inhibition in wild-type yeast *Saccharomyces cerevisiae* cells. *Biotechniques* 42, 158–162, 160, 162.
- Liu, O.W., Kelly, M.J.S., Chow, E.D., and Madhani, H.D. (2007b). Parallel β -helix proteins required for accurate capsule polysaccharide synthesis and virulence in the yeast *Cryptococcus neoformans*. *Eukaryot. Cell* 6, 630–640.
- Liu, O.W., Chun, C.D., Chow, E.D., Chen, C., Madhani, H.D., and Noble, S.M. (2008). Systematic genetic analysis of virulence in the human fungal pathogen *Cryptococcus neoformans*. *Cell* 135, 174–188.
- Loyse, A., Bicanic, T., and Jarvis, J.N. (2013). Combination antifungal therapy for cryptococcal meningitis. *N. Engl. J. Med.* 368, 2522–2523.
- Lozano, R., Naghavi, M., Foreman, K., Lim, S., Shibuya, K., Aboyans, V., Abraham, J., Adair, T., Aggarwal, R., Ahn, S.Y., et al. (2012). Global and regional mortality from 235 causes of death for 20 age groups in 1990 and 2010: a systematic analysis for the Global Burden of Disease Study 2010. *Lancet* 380, 2095–2128.
- Mandell, G.L., Bennett, J.E., and Dolin, R. (2010). *Mandell, Douglas, and Bennett's Principles and Practice of Infectious Diseases*, 7th edn (Philadelphia: Churchill Livingstone Elsevier).
- Meletiadi, J., Pournaras, S., Roilides, E., and Walsh, T.J. (2010). Defining fractional inhibitory concentration index cutoffs for additive interactions based on self-drug additive combinations, Monte Carlo simulation analysis, and in vitro-in vivo correlation data for antifungal drug combinations against *Aspergillus fumigatus*. *Antimicrob. Agents Chemother.* 54, 602–609.
- Morgan, D.O. (2007). *The Cell Cycle: Principles of Control* (London, UK: New Science Press).
- Nichols, R.J., Sen, S., Choo, Y.J., Beltrao, P., Zietek, M., Chaba, R., Lee, S., Kazmierczak, K.M., Lee, K.J., Wong, A., et al. (2011). Phenotypic landscape of a bacterial cell. *Cell* 144, 143–156.
- O'Meara, T.R., and Alspaugh, J.A. (2012). The *Cryptococcus neoformans* capsule: a sword and a shield. *Clin. Microbiol. Rev.* 25, 387–408.
- O'Meara, T.R., Norton, D., Price, M.S., Hay, C., Clements, M.F., Nichols, C.B., and Alspaugh, J.A. (2010). Interaction of *Cryptococcus neoformans* Rim101 and protein kinase A regulates capsule. *PLoS Pathog.* 6, e1000776.
- Park, B.J., Wannemuehler, K.A., Marston, B.J., Govender, N., Pappas, P.G., and Chiller, T.M. (2009). Estimation of the current global burden of cryptococcal meningitis among persons living with HIV/AIDS. *AIDS* 23, 525–530.

- Parsons, A.B., Brost, R.L., Ding, H., Li, Z., Zhang, C., Sheikh, B., Brown, G.W., Kane, P.M., Hughes, T.R., and Boone, C. (2004). Integration of chemical-genetic and genetic interaction data links bioactive compounds to cellular target pathways. *Nat. Biotechnol.* 22, 62–69.
- Parsons, A.B., Lopez, A., Givoni, I.E., Williams, D.E., Gray, C.A., Porter, J., Chua, G., Sopko, R., Brost, R.L., Ho, C.-H., et al. (2006). Exploring the mode-of-action of bioactive compounds by chemical-genetic profiling in yeast. *Cell* 126, 611–625.
- Pu, L., Amoscato, A.A., Bier, M.E., and Lazo, J.S. (2002). Dual G1 and G2 phase inhibition by a novel, selective Cdc25 inhibitor 6-chloro-7-[corrected][2-morpholin-4-ylethylamino]-quinoline-5,8-dione. *J. Biol. Chem.* 277, 46877–46885.
- Reese, A.J., and Doering, T.L. (2003). Cell wall α -1,3-glucan is required to anchor the *Cryptococcus neoformans* capsule. *Mol. Microbiol.* 50, 1401–1409.
- Roemer, T., Xu, D., Singh, S.B., Parish, C.A., Harris, G., Wang, H., Davies, J.E., and Bills, G.F. (2011). Confronting the challenges of natural product-based antifungal discovery. *Chem. Biol.* 18, 148–164.
- Taipale, M., Jarosz, D.F., and Lindquist, S. (2010). HSP90 at the hub of protein homeostasis: emerging mechanistic insights. *Nat. Rev. Mol. Cell Biol.* 11, 515–528.
- Vecchiarelli, A., Pericolini, E., Gabrielli, E., Kenno, S., Perito, S., Cenci, E., and Monari, C. (2013). Elucidating the immunological function of the *Cryptococcus neoformans* capsule. *Future Microbiol.* 8, 1107–1116.
- Whelan, W.L., and Kwon-Chung, K.J. (1986). Genetic complementation in *Cryptococcus neoformans*. *J. Bacteriol.* 166, 924–929.
- WHO (2011). Rapid Advice: diagnosis, prevention and management of cryptococcal disease in HIV-infected adults, adolescents, and children, W.H.O. Library, ed. (Geneva, Switzerland: WHO Press).
- Xu, D., Jiang, B., Ketela, T., Lemieux, S., Veillette, K., Martel, N., Davison, J., Sillaots, S., Trosok, S., Bachewich, C., et al. (2007). Genome-wide fitness test and mechanism-of-action studies of inhibitory compounds in *Candida albicans*. *PLoS Pathog.* 3, e92.
- Xu, D., Sillaots, S., Davison, J., Hu, W., Jiang, B., Kauffman, S., Martel, N., Ocampo, P., Oh, C., Trosok, S., et al. (2009). Chemical genetic profiling and characterization of small-molecule compounds that affect the biosynthesis of unsaturated fatty acids in *Candida albicans*. *J. Biol. Chem.* 284, 19754–19764.
- Zhai, B., Wu, C., Wang, L., Sachs, M.S., and Lin, X. (2012). The antidepressant sertraline provides a promising therapeutic option for neurotropic cryptococcal infections. *Antimicrob. Agents Chemother.* 56, 3758–3766.

A Widespread Glutamine-Sensing Mechanism in the Plant Kingdom

Vasuki-Ranjani Chellamuthu,^{1,3} Elena Ermilova,² Tatjana Lapina,² Jan Lüddecke,¹ Ekaterina Minaeva,² Christina Herrmann,¹ Marcus D. Hartmann,³ and Karl Forchhammer^{1,*}

¹Interfaculty Institute for Microbiology and Infection Medicine, University of Tübingen, Auf der Morgenstelle 28, 72076 Tübingen, Germany

²Lab Adaptation in Microorganisms, Biological Faculty, Saint-Petersburg State University, Universitetskaya em. 7/9, 199034 Saint-Petersburg, Russia

³Department of Protein Evolution, Max Planck Institute for Developmental Biology, Spemannstrasse 35, 72076 Tübingen, Germany

*Correspondence: karl.forchhammer@uni-tuebingen.de

<http://dx.doi.org/10.1016/j.cell.2014.10.015>

SUMMARY

Glutamine is the primary metabolite of nitrogen assimilation from inorganic nitrogen sources in microorganisms and plants. The ability to monitor cellular nitrogen status is pivotal for maintaining metabolic homeostasis and sustaining growth. The present study identifies a glutamine-sensing mechanism common in the entire plant kingdom except *Brassicaceae*. The plastid-localized P_{II} signaling protein controls, in a glutamine-dependent manner, the key enzyme of the ornithine synthesis pathway, *N*-acetyl-L-glutamate kinase (NAGK), that leads to arginine and polyamine formation. Crystal structures reveal that the plant-specific C-terminal extension of P_{II}, which we term the Q loop, forms a low-affinity glutamine-binding site. Glutamine binding alters P_{II} conformation, promoting interaction and activation of NAGK. The binding motif is highly conserved in plants except *Brassicaceae*. A functional Q loop restores glutamine sensing in a recombinant *Arabidopsis thaliana* P_{II} protein, demonstrating the modular concept of the glutamine-sensing mechanism adopted by P_{II} proteins during the evolution of plant chloroplasts.

INTRODUCTION

Glutamine is the primary product of nitrogen assimilation from inorganic nitrogen sources and a central metabolite in nitrogen metabolism in plants. The regulation of nitrogen metabolism and the maintenance of a proper carbon-nitrogen balance in plants are extremely complex and interwoven at multiple levels. How plants sense nitrogen status is an intriguing issue of current research (e.g., Nunes-Nesi et al., 2010). Glutamine is an important nitrogen status reporter in many bacteria (Forchhammer, 2007) but, in plants, glutamine sensing is poorly understood. In prokaryotes, P_{II} signaling proteins (in the following termed P_{II} proteins) have been identified as central signal-integrating molecules coordinating nitrogen assimilatory reactions in response to the general metabolic state of cells. The P_{II} protein family repre-

sents one of the most abundant families of signaling proteins in nature, occurring in bacteria, Archaea, and plants (Chellamuthu et al., 2013; Leigh and Dodsworth, 2007; Sant'Anna et al., 2009). P_{II} proteins are able to sense and integrate signals from central metabolism, in particular 2-oxoglutarate (2-OG), an indicator of C-N balance, as well as the energy status via competitive ADP/ATP binding (Fokina et al., 2010; Jiang and Ninfa, 2009b; Zeth et al., 2014).

The concentration of these metabolite effectors is sensed by binding to an ensemble of intercommunicating sites in the trimeric P_{II} proteins, where the actual effector molecule binding status is translated into distinct P_{II} conformations. The T loop, a highly flexible loop segment protruding from each subunit, plays a key role in ligand binding and receptor interactions (Forchhammer, 2008; Huergo et al., 2013). Depending on the conformation of the T loop, P_{II} proteins can bind to various receptors, thereby exerting control at all levels of metabolic regulation (transport activity, metabolic reactions, gene expression). This basic mode of P_{II} function, based on direct sensing of effector molecules, is conserved in almost all P_{II} proteins.

A second, phylogenetically diverse regulatory layer may be superposed on P_{II} signaling: covalent modification of apical residues of the T loop allows the integration of additional signals. In proteobacteria, P_{II} proteins can be covalently modified by uridylylation at T loop residue Tyr51. The modifying enzyme GlnD, a bifunctional uridylyltransferase/uridylyl-removing enzyme (EC 2.7.7.59), responds to the cellular glutamine level such that low glutamine levels trigger the P_{II}-uridylylation reaction (Adler et al., 1975; Ninfa and Jiang, 2005). This results in Tyr51-uridylylated P_{II}, which is, for instance, not able to interact with the ammonium channel AmtB (Radchenko and Merrick, 2011). Conversely, increasing glutamine levels cause P_{II}-deuridylylation (Adler et al., 1975), allowing the interaction of P_{II} with various targets (Huergo et al., 2013). As a consequence, glutamine-dependent uridylylation/deuridylylation of P_{II} is the prevailing signaling mechanism of the *Escherichia coli* P_{II} system (Forchhammer, 2007; Ninfa and Atkinson, 2000). In analogy to proteobacterial P_{II}-uridylylation, P_{II} in actinobacteria is adenylated at Tyr51 in a glutamine-dependent manner (Hesketh et al., 2002). In cyanobacteria, the neighboring residue Ser49 is phosphorylated in response to nitrogen starvation (Forchhammer and Tandeau de Marsac, 1994), but the detailed molecular signals controlling the kinase activity are still unknown. In many other organisms, this second regulatory layer of covalent modification of the T

loop is apparently missing, as in *Bacillus* (Heinrich et al., 2006), Archaea (Leigh and Dodsworth, 2007), and plant P_{II} proteins (Smith et al., 2004; Uhrig et al., 2009).

Phylogenetic analysis has inferred that plant P_{II} proteins are of cyanobacterial origin and have been conserved during the evolution of the *Chloroplastida* (plant kingdom, comprising green algae and land plants) from the ancestral cyanobacterial endosymbiont to higher plants. By contrast, P_{II} signaling has been lost in some red algae and in the *Chromalveolata* (Chellamuthu et al., 2013). The only plant P_{II} protein that has been studied biochemically in depth is that of *Arabidopsis thaliana* (Mizuno et al., 2007a; Smith et al., 2003). Physiological studies indicate that plant P_{II} proteins are part of a complex signal-transduction network mediating nitrogen regulation (Hsieh et al., 1998; Nunes-Nesi et al., 2010). P_{II} mutants of *A. thaliana* are affected in the synthesis of compounds derived from the ornithine/arginine synthesis pathway (Ferrario-Méry et al., 2006), overaccumulate carbon metabolites (Ferrario-Méry et al., 2005), and show impaired control of nitrite uptake into the chloroplast (Ferrario-Méry et al., 2008). Furthermore, P_{II} seems to be required for proper seed development (Uhrig et al., 2009). Such tissue-specific roles of P_{II} signaling together with the tightly intertwined complex network of signal-transduction cascades (Nunes-Nesi et al., 2010) complicate the elucidation of P_{II} -specific functions in higher plants. To study fundamental problems in plant chloroplasts, the simple unicellular green alga *Chlamydomonas reinhardtii* has proved an excellent model organism (Harris, 2001). For this reason, we started studying P_{II} signaling in *C. reinhardtii* (Ermilova et al., 2013).

Like its higher plant homologs, *C. reinhardtii* P_{II} (CrP_{II}) is localized in the chloroplast. Residues involved in effector molecule binding are conserved, which suggests that CrP_{II} functions as a signaling protein as well. The chloroplast-localized enzyme *N*-acetyl-L-glutamate kinase (NAGK), which catalyzes the committed step in the ornithine/arginine biosynthesis pathway, is an important P_{II} -interacting protein in cyanobacteria and higher plants (Ferrario-Méry et al., 2006; Heinrich et al., 2004; Sugiyama et al., 2004). The structures of the P_{II} -NAGK complex from the cyanobacterium *Synechococcus elongatus* and from *A. thaliana* have been solved and are highly similar (Llácer et al., 2007; Mizuno et al., 2007b). Beez et al. (2009) have shown that the P_{II} -NAGK proteins from *S. elongatus* and *A. thaliana* functionally complement each other in vitro, highlighting striking functional conservation. Binding of P_{II} enhances NAGK activity in the presence of the feedback inhibitor arginine, an effect that can be antagonized by the P_{II} -effector molecule 2-OG. When P_{II} binds to NAGK, the T loops of P_{II} must fold in a tightly flexed structure to insert into the interdomain crevice of the adjacent NAGK subunits (Llácer et al., 2007; Mizuno et al., 2007b).

An unsolved question in the evolution of plant P_{II} proteins concerns the appearance of an elongated C terminus. Whereas all bacterial P_{II} proteins have an almost invariable C terminus, the P_{II} proteins from the plant kingdom contain a C-terminal extension of 13–19 amino acids with a conserved motif. The function of the C-terminal extension is so far unknown. In the present work, we studied the structural and biochemical properties of CrP_{II} . We discovered that the C-terminal extension of CrP_{II} forms a small loop structure (termed the Q loop) that binds glutamine and is required for glutamine-dependent complex formation

with NAGK. The same glutamine dependence was observed for recombinant *Physcomitrella* and *Oryza* P_{II} proteins, which implies that plant P_{II} proteins function as glutamine sensors via the C-terminal Q loop extension.

RESULTS

Basic Properties of *C. reinhardtii* NAGK, the Putative P_{II} Target

The predicted full-length NAGK polypeptide encoded by the *C. reinhardtii* AGK1 cDNA consists of 340 amino acids with a calculated molecular mass of 35,971 Da. The N terminus contains a predicted chloroplast transit peptide (residues 1–42), which suggests that, like its higher plant homologs, *C. reinhardtii* NAGK ($CrNAGK$) protein resides in the chloroplast with a molecular mass of 31,474 Da. A previous survey showed that all NAGK proteins from cyanobacteria and plants (*Chloroplastida*), and from those red algae that contain P_{II} proteins, share the signature residues involved in NAGK- P_{II} interaction (Chellamuthu et al., 2013; an alignment is shown in Figure S1 available online), which strongly implies that they all are targets of P_{II} signaling. To study the biochemical properties and to characterize the putative interaction between CrP_{II} and $CrNAGK$, a recombinant $CrNAGK$ protein was generated (starting from residue Met44) with its N terminus fused to a His₆ tag. Analysis of the purified $CrNAGK$ protein by gel filtration revealed that most of the protein eluted in the expected hexameric state, whereas a smaller fraction eluted with an apparent higher mass. By comparison, *A. thaliana* NAGK ($AtNAGK$) eluted as a single peak, corresponding to the hexameric enzyme (Figure S2). The kinetic constants of $CrNAGK$ showed a K_m value for *N*-acetyl glutamate of 7.77 mM and a k_{cat} of 56.76 s⁻¹ (calculated with NAGK hexamer). This K_m value is 9-fold higher than that of $AtNAGK$ but very similar to that of the cyanobacterial *S. elongatus* NAGK ($SeNAGK$) (K_m [NAG] of 7.4 mM), whereas the k_{cat} value is intermediate between that of $AtNAGK$ (126 s⁻¹) and $SeNAGK$ (13 s⁻¹). Feedback inhibition by arginine occurred with a half-maximal inhibitory concentration (IC₅₀) of 0.11 mM (Figure 1B), which is intermediate between that of cyanobacterial $SeNAGK$ (0.02 mM) and plant $AtNAGK$ (1 mM).

Glutamine-Dependent Interaction between CrP_{II} and NAGK

To test the effect of CrP_{II} on the catalytic activity of $CrNAGK$, we assayed NAGK using a coupled ATPase assay (Beez et al., 2009). In a control, ATPase activity of CrP_{II} in the absence of NAGK was below the detection limit of the assay. The previously reported ATP hydrolysis activity of P_{II} proteins is not catalytic, but needs a large excess of P_{II} over ATP (Radchenko et al., 2013).

Determining the relief from arginine inhibition by P_{II} -NAGK complex formation represents the most sensitive assay for P_{II} -NAGK interaction. In arginine-inhibition assays with $CrNAGK$ in the presence or absence of CrP_{II} (Figure 1A), the presence of CrP_{II} surprisingly did not change arginine inhibition of $CrNAGK$, nor did it change its catalytic activity. As a possible suitable positive control of CrP_{II} function, we tested the effect of CrP_{II} on $AtNAGK$, which we previously showed to functionally interact

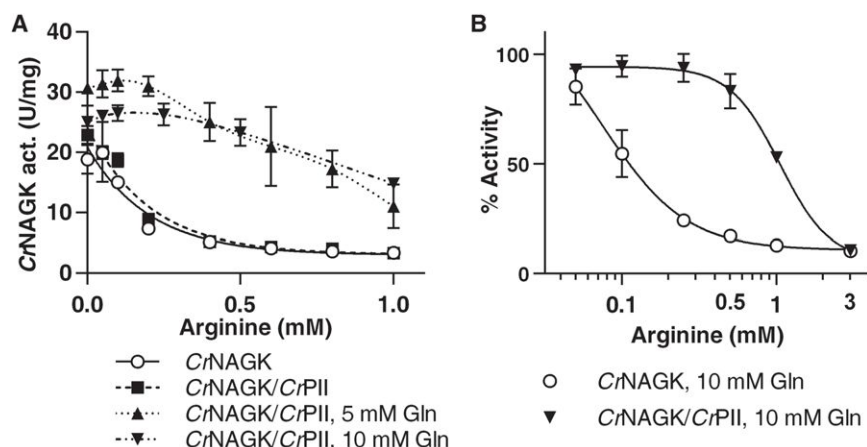


Figure 1. Arginine Inhibition of CrNAGK Activity in the Presence of CrP_{II}

(A) Arginine inhibition of CrNAGK in the absence or presence of CrP_{II}, without or with 5 or 10 mM glutamine, as indicated.

(B) Arginine inhibition of CrNAGK in the presence of 10 mM glutamine or in the presence of CrP_{II} and 10 mM glutamine. Data were fitted to a sigmoidal dose-response curve, yielding an EC₅₀ (arginine) of 0.11 ± 0.04 mM and 1.05 ± 0.15 mM for free or CrP_{II}-interacting CrNAGK, respectively.

When exceeding the size of the data point symbols, standard deviation from triplicate measurements is indicated by error bars.

not only with its own P_{II} protein but also with P_{II} from the cyanobacterium *S. elongatus* (Beez et al., 2009). However, CrP_{II} did not affect the arginine response of AtNAGK (data not shown). We therefore suspected that an additional component may be required for CrP_{II} interaction with NAGK. We added amino acids and metabolites from core metabolism (glutamate, glutamine, aspartate, asparagine, alanine, serine, glycine, acetyl-CoA, and tricarboxylic acid cycle intermediates) to the arginine-inhibition assays of CrNAGK-CrP_{II}. Of the various compounds tested, glutamine indeed rescued the activity of CrP_{II}; in the presence of 5 or 10 mM glutamine, the inhibitory effect of arginine was strongly antagonized (Figure 1A). In the presence of P_{II} and 10 mM glutamine, the IC₅₀ of arginine for CrNAGK increased from 0.1 mM to 1 mM (Figure 1B). The highest difference in activity between CrNAGK with or without CrP_{II} was with 0.4 mM arginine. Therefore, we titrated the effect of glutamine with 0.4 mM arginine. Glutamine activated arginine-inhibited CrNAGK in the presence of CrP_{II} in a concentration-dependent manner (Figure 2A), whereas glutamine had no effect in the absence of CrP_{II}. The half-maximal effective concentration (EC₅₀) of glutamine for activation of NAGK by CrP_{II} was 2.4 ± 0.8 mM. To reveal whether the glutamine dependence of CrP_{II}-NAGK activation is a property of CrP_{II} or of CrNAGK, heterologous assays were carried out using AtNAGK. CrP_{II} had no effect on AtNAGK in the absence of glutamine. By contrast, arginine inhibition was relieved by glutamine in a concentration-dependent manner (Figure 2B). The response was similar to the glutamine-dependent activation of CrNAGK by CrP_{II}, with a glutamine EC₅₀ of 4.6 ± 2.4 mM, which showed that the glutamine dependence is a property of CrP_{II}.

To determine the response of CrP_{II}-CrNAGK interaction toward 2-OG, we titrated CrP_{II} and CrNAGK (5 CrP_{II} trimers:1 CrNAGK hexamer) in the presence of 10 mM glutamine and 0.4 mM arginine with increasing concentrations of 2-OG. 2-OG tuned down NAGK activity with an IC₅₀ of 1.18 ± 0.02 mM (Figure 2C). An almost identical response toward 2-OG was obtained for the heterologous assay with CrP_{II} and AtNAGK (Figure 2D; IC₅₀ of 1.26 ± 0.96 mM). The catalytic constants of the CrP_{II}-CrNAGK complex determined at 5 mM glutamine showed that CrP_{II} does not enhance the V_{max} (54 s^{-1}) of the NAGK reaction but lowers the K_m for NAG about 2-fold (3.9 ± 0.4 mM as compared to 7.8 ± 0.8 mM).

Analysis of CrP_{II}-NAGK Complex Formation

To confirm that the glutamine-dependent regulatory effect of CrP_{II} on NAGK is due to direct interaction, we analyzed complex formation using surface plasmon resonance (SPR) spectroscopy. The N-terminally His-tagged CrNAGK protein was immobilized on an Ni-NTA sensor chip and probed with CrP_{II} together with effector molecules. CrP_{II} bound to CrNAGK in a glutamine-dependent manner; with 5 mM glutamine, complex formation appeared to be saturated (Figure 3A). This result confirmed that complex formation between CrP_{II} and CrNAGK is strictly glutamine dependent. Furthermore, complex formation of CrP_{II} with CrNAGK strictly required Mg-ATP and was not supported by ADP (Figure 3B). Finally, we analyzed the effect of 2-OG on CrP_{II}-NAGK complex formation. Regardless of the 2-OG concentration added to the assay mixture (containing 1 μ M CrP_{II} and 5 mM glutamine), no antagonistic effect on complex formation was observed (Figure 3C), although these 2-OG concentrations prevented activation of NAGK (Figure 2C). This result resembles the case of *A. thaliana* P_{II}-NAGK interaction, in which the 2-OG concentrations required to inhibit complex formation are three orders of magnitude higher than those required to impair NAGK activation by P_{II} (Beez et al., 2009). Similarly, in *E. coli*, the regulation of NtrB activity by P_{II}, but not its binding, is controlled by 2-OG (Jiang and Ninfa, 2009a). Thus, the inhibitory effect of 2-OG on P_{II}-mediated activation of NAGK in *C. reinhardtii* appears to occur postbinding.

Structural Characterization of CrP_{II}

The EC₅₀ of glutamine for stimulating CrP_{II}-NAGK interaction was in the millimolar range, which indicated a low affinity of CrP_{II} for glutamine, outside the sensitivity range of direct ligand-binding assays. Titration of glutamine to CrP_{II} by isothermal titration calorimetry (ITC) yielded only marginal calorimetric signals. To define the role of glutamine and identify its putative binding site in CrP_{II}, we used crystallographic analysis. First, we set up CrP_{II} crystallization trials in the presence of different effector molecules. From these trials, we obtained well-diffracting crystals that grew in the presence of Mg-ATP and glutamine, for which we could solve the structure by molecular replacement using the SeP_{II} structure (Protein Data Bank [PDB] ID code) 2XZW as a search model. The resulting CrP_{II} structure has the expected

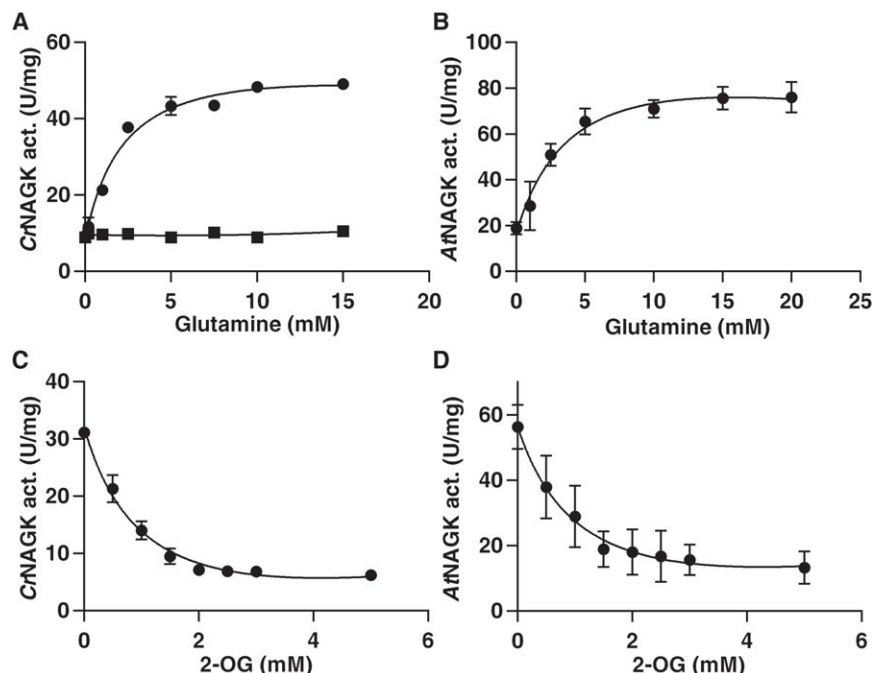


Figure 2. Effect of Glutamine or 2-OG on CrP_{II} Activation of $CrNAGK$ and $AtNAGK$

$CrNAGK$ assays contained 0.4 mM arginine; $AtNAGK$ assays contained 3 mM arginine.

(A) Glutamine-dependent activation of $CrNAGK$ by CrP_{II} (circles) or by $CrP_{II}-\Delta C$ (squares).

(B) Glutamine-dependent activation of $AtNAGK$ by CrP_{II} .

(C and D) Antagonistic effect of 2-OG on $CrP_{II}/10$ mM glutamine-mediated activation of (C) $CrNAGK$ or (D) $AtNAGK$.

The standard deviation from triplicate measurements is indicated by error bars.

trimeric assembly and superimposes (with a root-mean-square deviation [rmsd] of 0.85 Å over 271 aligned $C\alpha$ positions) closely on the SeP_{II} structure 2XZW. However, neither Mg-ATP nor glutamine was identified in the structure, and both the T loop and the whole C-terminal extension were disordered (Figure 4A). Therefore, we then tried to bind ligands to the protein by soaking the apo crystals in solutions containing different combinations of effector molecules. This was successful for the combination of Mg-ATP and 2-OG, which resulted in a crystal structure with asymmetrically occupied effector sites—one site occupied by Mg-ATP and 2-OG in full occupancy, the second site occupied by Mg-ATP and 2-OG in half-occupancy, and the third site empty (Figure 4B). Half-occupancy denotes that, on average, throughout the crystal, only every second P_{II} trimer has a molecule bound in the respective binding site, which indicates an interaction weaker than binding with full occupancy. The structure superimposes with an rmsd of 0.47 Å on the apo structure, and both the T loop and C-terminal extension are unstructured in all subunits. The binding mode of the effector molecules is essentially the same as reported for various bacterial and archaeal P_{II} proteins (Fokina et al., 2010; Maier et al., 2011; Truan et al., 2010). The asymmetrical filling of the sites suggests that binding of the effector molecules is anticooperative, as reported previously for the *E. coli* (Jiang and Ninfa, 2009a) and *S. elongatus* P_{II} protein (Fokina et al., 2010; Zeth et al., 2014).

Identification of the Glutamine-Binding Site in the P_{II} -NAGK Complex Structure

Crystallization trials were carried out with CrP_{II} -NAGK complexes in the presence of glutamine and different combinations of effector molecules. Because initial crystallization trials with the homologous $CrNAGK$ were not fruitful, we set up trials with the heterologous $AtNAGK$, because the glutamine response of

the CrP_{II} - $AtNAGK$ complex was almost identical to that of the homologous CrP_{II} - $CrNAGK$ complex. This setup rewarded us with crystals of the whole complex that finally disclosed the binding mode of glutamine. Glutamine was localized to the C-terminal extension of CrP_{II} . Interestingly, these CrP_{II} - $AtNAGK$ crystals have the same crystal packing as those of the AtP_{II} - $AtNAGK$ structure (PDB ID code) 2RD5, so that the structure could

be solved directly on the basis of the 2RD5 coordinates (see Extended Experimental Procedures). Consequently, the complex is almost identical to the AtP_{II} - $AtNAGK$ complex and very similar to the SeP_{II} -NAGK complex, with a hexameric NAGK toroid sandwiched between two P_{II} trimers (Figure 5A). As described for the *A. thaliana* P_{II} -NAGK complex (Mizuno et al., 2007b), the NAGK subunits are in an open conformation on one face of the toroid and in a closed conformation on the other face. The ligand-binding sites of the closed subunits are occupied with arginine, NAG, and Mg-ADP, whereas the subunits in open conformation are occupied with arginine and the reaction product NAG-phosphate but not with a nucleotide. P_{II} is in complex with the desired glutamine and Mg-ATP, although no ATP had been added to the crystallization solution (Figure 5B). The presence of ATP, presumably originating from a contamination, further strengthens the conclusion that CrP_{II} can only bind $AtNAGK$ in the Mg-ATP complex.

Because most residues important for complex formation are conserved between CrP_{II} and AtP_{II} , it is not surprising that the overall structure and in particular the conformation of the T loop of the CrP_{II} - $AtNAGK$ and AtP_{II} - $AtNAGK$ (2RD5) complexes are nearly identical. A superposition of P_{II} trimers from both complexes yields an rmsd of 0.79 Å over 372 aligned $C\alpha$ positions, which supports the significance of the CrP_{II} conformation in the heterologous complex.

The remarkable and important difference between AtP_{II} and CrP_{II} in complex with NAGK concerns the C-terminal extension. In glutamine-bound CrP_{II} , the extension forms two additional short α -helical segments connected by a small loop, from Lys131 to Gly135 (Figure 5C). This extension constitutes a large portion of the glutamine-binding site. Glutamine mainly forms hydrogen bonds to main-chain atoms of this extension and to Arg43, Gly44, and Val82 (Figure 5D). Additional hydrogen bonds

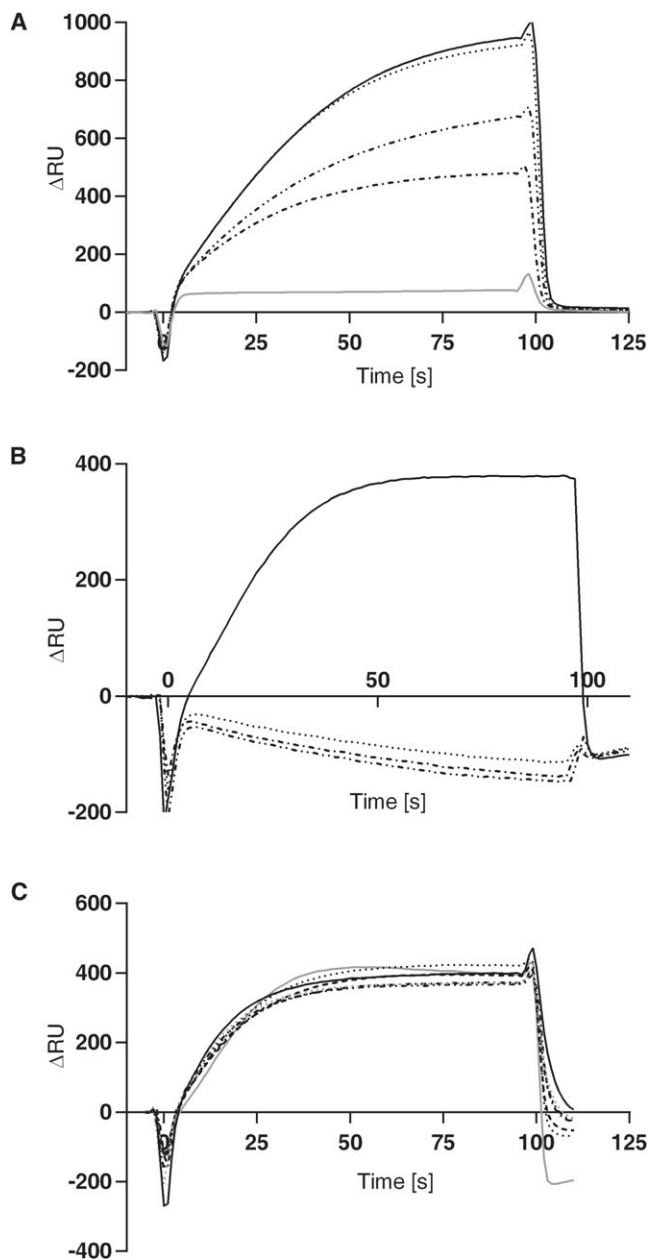


Figure 3. CrP_{II}-CrNAGK Complex Formation Analyzed by SPR Spectroscopy

CrP_{II} with effectors was injected on immobilized CrNAGK in flow cell 2; the response difference between FC2 and the control FC1 is shown.

(A) Glutamine-dependent binding of CrP_{II} to CrNAGK at 0 (gray line), 1 mM (dot-dashed line), 2.5 mM (dot-dot-dashed line), 5 mM (dotted line), and 7.5 mM (black line) glutamine.

(B) ATP-dependent complex formation at 5 mM glutamine. Lines: black, 1 mM ATP/2 mM MgCl₂; dotted, 1 mM ADP; dash-dotted, 1 mM ADP/2 mM MgCl₂; dot-dot-dashed, 1 mM ADP/2 mM MgCl₂/5 mM glutamine.

(C) Influence of 2-OG on CrP_{II}-CrNAGK interaction at 5 mM glutamine. Gray line, no 2-OG; other lines, 0.5, 1, 3, 5, 10, or 20 mM 2-OG.

are formed with the side chains of Glu130 and Gln87. The structure of the unique C-terminal loop (from Lys131 to Gly135), which we term the Q loop, is further stabilized by contacts with a region from the basal part of the T loop, which also takes part in effector molecule binding (Figure 6A). Two hydrogen bonds are formed between the Q loop and T loop residues Gln56 and Gly57. Of particular interest is Gln56 (corresponding to Gln39 in bacterial P_{II} proteins), as it is a key residue for the structure of the T loop, switching between different P_{II} structural states (Huergo et al., 2013; Zeth et al., 2014). It seems therefore that its interactions with the Q loop help to bring the T loop into its flexed conformation, required for complex formation with NAGK.

Functional Implications of Glutamine Binding

Although the crystal structure reveals the binding site of glutamine, it does not mechanistically explain why glutamine is essential for CrP_{II} complex formation. Either the C-terminal segment acts as an antideterminant, preventing binding to NAGK, or the Q loop in the glutamine-bound state supports complex formation. To distinguish between these possibilities, the C-terminal segment of CrP_{II} was deleted, shortening the protein to the canonical length of bacterial P_{II} proteins. The recombinant truncated CrP_{II} protein (CrP_{II}-ΔC) has the expected size of a trimeric P_{II} protein (Figures S3A and S3B). Furthermore, the truncated protein retained the core function of P_{II} proteins, as deduced from 2-OG-binding assays using ITC. The isotherms could be fitted to a one-site binding model with a K_d of 90 μM, as compared to 39 μM for the full-length protein (Figures S3C and S3D). Interaction of CrP_{II}-ΔC with NAGK was tested in enzyme assays (Figure 2A) as well as in SPR experiments (Figure S3E). In no assays was interaction detected, neither in the presence nor absence of glutamine, which indicated that the glutamine-binding C-terminal segment is actively required for CrP_{II} to bind to NAGK in a glutamine-dependent manner.

Glutamine Sensing Is a General Property of Plant P_{II} Proteins

Sequence alignment of plant P_{II} proteins (Figure 6C; an extended list is in Figure S4) reveals that the C-terminal residues of the Q loop are part of a conserved motif present in all plant sequences that we analyzed, except the *Brassicaceae* family, to which *A. thaliana* belongs. The Q loop in members of the *Brassicaceae* family exhibits a deletion of three amino acids, which might help the Q loop to fold in a glutamine-independent manner (Figure 6B). To find out whether the missing glutamine response of *A. thaliana* P_{II} is a consequence of this genetic trait, we exchanged the 19 amino acid Q loop extension of CrP_{II} (see Figure 6C) with the 15 amino acid C terminus of AtP_{II}, and vice versa, to obtain chimeric CrP_{II} with the *A. thaliana* C terminus (CrP_{II}/AtQ) and chimeric AtP_{II} with the *C. reinhardtii* Q loop terminus (AtP_{II}/CrQ). The CrP_{II}/AtQ chimera only very weakly activated CrNAGK (Figure 7A), which indicated that the C terminus of *A. thaliana* P_{II} cannot efficiently replace that of *C. reinhardtii*, and that this replacement completely abrogated the glutamine response. Conversely, the AtP_{II}/CrQ chimera was functional and, moreover, showed partial glutamine dependence (Figure 7B). Therefore, the inability of *A. thaliana* P_{II} to respond to glutamine was indeed caused by the degenerated C terminus. The conserved

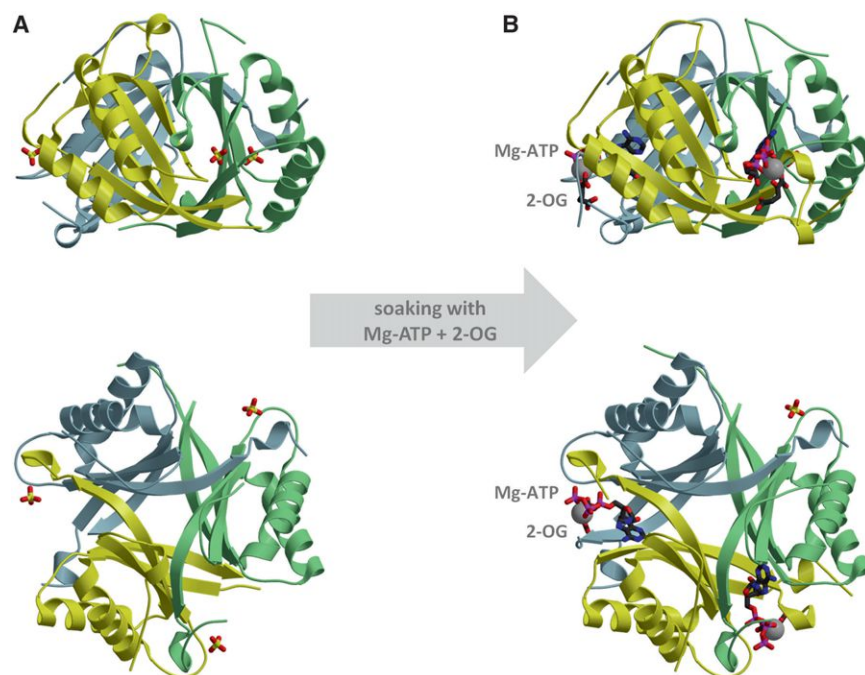


Figure 4. Side and Top Views of the Trimer Structure of CrP_{II} in the Apo State and with Bound Mg-ATP and 2-OG

In the apo structure (A), sulfate ions from the reservoir solution are bound in place of the ATP γ -phosphate in all three monomers. Upon soaking these crystals with Mg-ATP and 2-OG (B), Mg-ATP is bound with full occupancy in two monomers; 2-OG is bound fully in one monomer, with half-occupancy in the other. The third monomer remains in the sulfate-bound apo state. The T loop and the C-terminal extension including the Q loop are unstructured in all states. Data collection and refinement statistics for the structures are shown in Table S1.

Q loop C terminus of plant P_{II} proteins not belonging to the *Brassicaceae* strongly implies that glutamine sensing might be a general feature of P_{II} function in plants. To test this assumption, we prepared recombinant P_{II} proteins from two very diverse plants, the moss *Physcomitrella patens* and the rice *Oryza sativa*. To test the glutamine response, these P_{II} proteins were added to arginine-feedback-inhibited AtNAGK with increasing concentrations of glutamine. Both P_{II} proteins were able to release AtNAGK from arginine-feedback inhibition in a glutamine-dependent manner: *Physcomitrella* P_{II} with an EC₅₀ of 6.6 mM, and *Oryza* P_{II} with an EC₅₀ of 9.2 mM (Figure 7C). This glutamine response is similar to that of CrP_{II} (Figure 2B), which confirmed that glutamine sensing is a general property of plant P_{II} proteins.

DISCUSSION

Maintenance of a proper carbon-nitrogen balance is pivotal for plant growth and development. A functional understanding of the underlying regulatory mechanisms is fundamental to improvement of crop yield and resistance (Nunes-Nesi et al., 2010; Lea and Mifflin, 2010). In higher plants, the signal-transduction network associated with nitrogen assimilation is extremely complex, integrating signals from hormones, nitrate, sugars, organic acids, and amino acids (Foyer et al., 2003; Nunes-Nesi et al., 2010; Zheng, 2009). Glutamine, as the primary product of ammonium assimilation, is of particular importance, and is one of the most abundant amino acids of light-grown plants (Fritz et al., 2006). Ammonium is provided by nitrate reduction, photorespiration, and amino acid recycling, and is mainly assimilated in the chloroplast by glutamine synthetase (Keys, 2006; Lea and Mifflin, 2010). Glutamine donates nitrogen groups, directly or via glutamate, for essentially all nitrogenous cell compounds.

highly conserved from cyanobacteria to higher plants: P_{II} enhances the catalytic activity of arginine-feedback-inhibited NAGK when P_{II} is ATP bound and not in complex with 2-OG and, conversely, ADP or 2-OG antagonizes productive interaction between P_{II} and NAGK (Beez et al., 2009; Mizuno et al., 2007b). From the conservation of critical amino acids in P_{II} and NAGK proteins involved in complex formation, we assumed that this interaction would be conserved in green algae, the phylogenetic predecessors of higher plants. Surprisingly, the *Chlamydomonas* P_{II} protein required glutamine in millimolar concentrations to interact with NAGK. Moreover, P_{II} proteins from the moss *Physcomitrella* and from the rice *Oryza* both showed a glutamine requirement similar to *Chlamydomonas* P_{II}. By contrast, *A. thaliana* P_{II} acted independent of glutamine. The slight truncation of the AtP_{II} C-terminal extension turned out to be responsible for the divergent property.

To mechanistically explain why AtP_{II} binds to NAGK independent of glutamine in contrast to the other P_{II} proteins requires an in-depth comparison of the structures. The C-terminal extension of CrP_{II} forms a small helix-Q loop-helix structure, which wraps around the bound glutamine molecule. The bound glutamine acts like a pillar that puts the Q loop in place through its carboxyl group and holds the Q loop through main-chain interactions. Backbone interactions from the Q loop with Gln56 and Gly57 from the base of the T loop stabilize the highly flexible T loop and Q loop structures. In agreement, in crystals of CrP_{II} not ligated to NAGK, these structures are disordered, which suggests that the flexible T loop and Q loop become structured upon NAGK binding. This agrees with the fact that NAGK arranges major contacts to the T loop of P_{II}. A CrP_{II} variant with the C terminus truncated to the length of bacterial P_{II} proteins was unable to interact with NAGK, although it appeared to be an intact P_{II} protein with respect to ternary structure and 2-OG

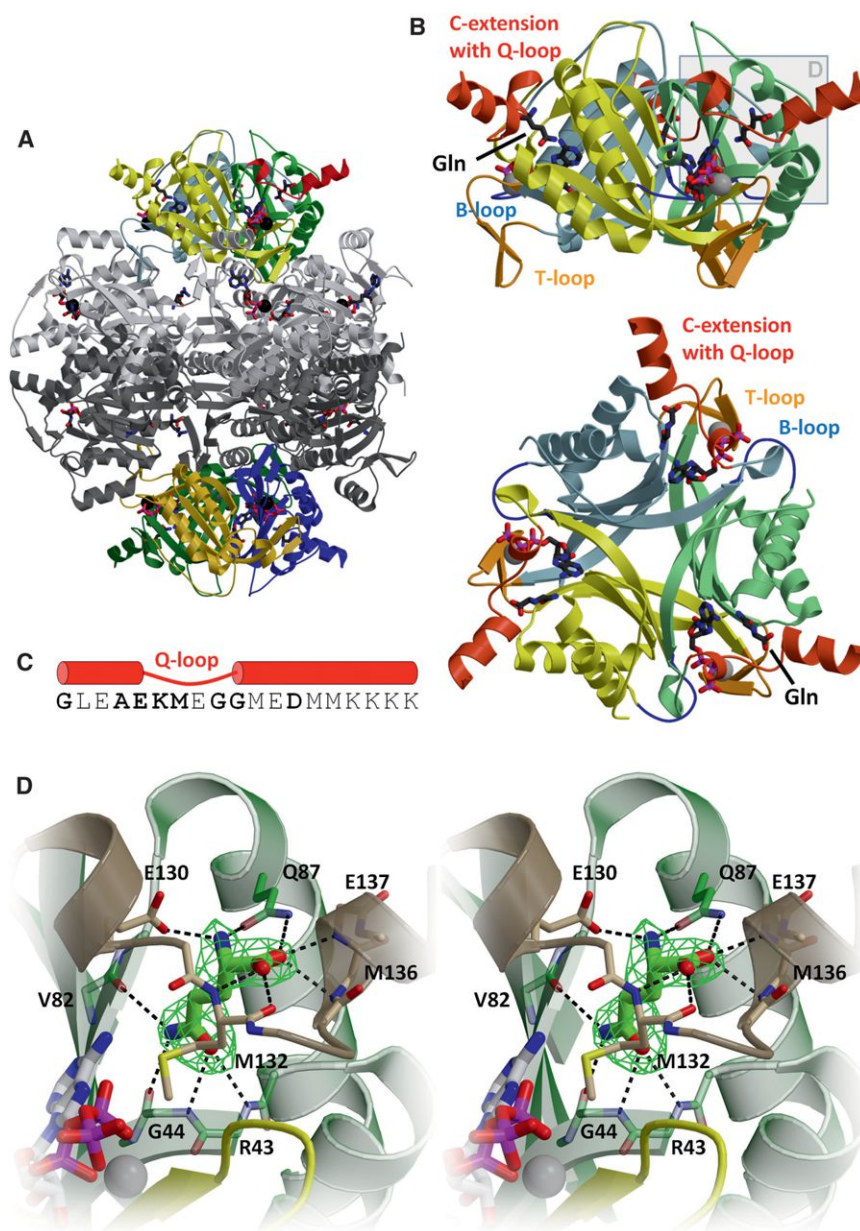


Figure 5. Glutamine-Bound CrP_{II} in Complex with AtNAGK

(A) P_{II} monomers are individually colored; NAGK is in two shades of gray. In the P_{II} trimer on top, one C-terminal extension is highlighted in red.

(B) Side and top views of the P_{II} trimer from the complex. The T loop, B loop, and whole C-terminal extension are individually colored. The whole C-terminal extension, consisting of two α helices connected by the Q loop, constitutes a large portion of the glutamine-binding site.

(C) The sequence of the C-terminal extension, highlighting the consensus motif.

(D) A stereo close-up of the glutamine-binding site reveals a sophisticated hydrogen-bonding network. Electron density for glutamine is shown as an F_O - F_C omit map contoured at 4 σ . Data collection and refinement statistics of the CrP_{II}-AtNAGK complex are shown in Table S1.

binding. This suggests that the T loop in the absence of the Q loop cannot adopt the structure that binds NAGK. The interactions between the Q loop and the T loop are apparently necessary to stabilize the structure. The same stabilizing interactions between the C terminus and the T loop are found in the *A. thaliana* P_{II}-NAGK complex (see Figure 6B). However, in this case, no binding of glutamine is required. Owing to the deletion of three amino acids at the Q loop, an aspartate residue (Asp129) projects into the glutamine-binding pocket and a hydrogen-bonding network with water molecules fills the cavity. A key residue for determining the structure of the T loop in CrP_{II} is Gln56 (corresponding to Gln50 in AtP_{II} and Gln39 in *S. elongatus* P_{II}). The importance of this residue for stabilizing T loop conformations was originally recognized in *E. coli* P_{II} protein GlnB (Jiang

et al., 1997). In bacterial P_{II} proteins, different Gln39 contacts switch the T loop in various states (Huergo et al., 2013; Zeth et al., 2014). In both CrP_{II} and AtP_{II} complexes with AtNAGK, the corresponding glutamine side chain forms a bridge between the γ -phosphate oxygen of the bound ATP molecule and the backbone Met132 (CrP_{II}) (Met126 in AtP_{II}), thereby stabilizing the tightly flexed T loop conformation. As observed here for the CrP_{II}-AtNAGK structure, all P_{II} sites of the *A. thaliana* P_{II}-NAGK structure are occupied by ATP, although no ATP was added to the crystallization solution and must have been acquired from contaminating ATP in the ADP solution (Mizuno et al., 2007b). This demonstrates the strict ATP requirement for complex formation. This can be explained if we assume that the Gln56 interaction between ATP and the C-terminal backbone is necessary to fix the T loop in the flexed conformation necessary for insertion

into the crevice of NAGK. In cyanobacteria, the C-terminal Q loop extension does not exist but, nevertheless, P_{II} avidly interacts with NAGK. Moreover, this interaction does not require ATP, which indicates that in this case the T loop can fold into the NAGK-fitting structure without engaging the Gln39-ATP contact. Indeed, in the *S. elongatus* P_{II}-NAGK structure, which is devoid of ATP, Gln39 is not involved in any contacts (Llácer et al., 2007).

Of 55 examined P_{II} C-terminal sequences of plants including green algae (Figure S4), only the 5 representatives of the *Brassicaceae* family have the deletion found in *Arabidopsis*, and of the 50 remaining cases, only 3 have significant alterations in the Q loop consensus motif. From this almost invariant conservation, we deduced that the formation of the P_{II}-NAGK complex in

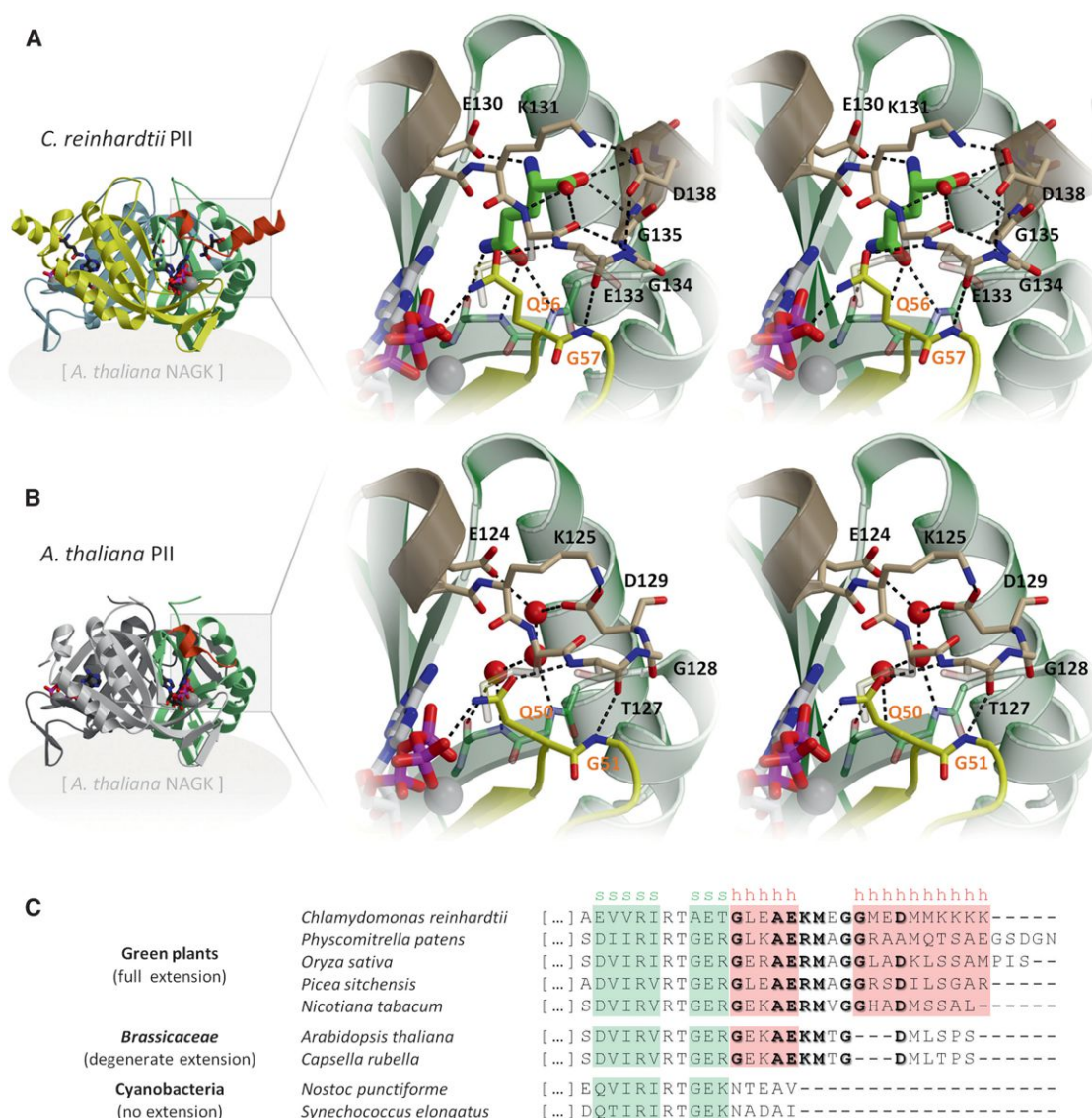


Figure 6. Detailed View of the Interactions between the Q Loop and T Loop in PII from *C. reinhardtii* and *A. thaliana*

(A and B) *C. reinhardtii* (A) and *A. thaliana* (B). In both complexes, the structured Q loop forms the same stabilizing hydrogen bonds with the γ -phosphate-coordinating glutamine in the T loop (Q56 in CrPII, Q50 in AtPII) and the subsequent glycine (G57/G51). Structurally, the twin glycine residues (G134 and G135 in CrPII) of the Q loop provide space for the glutamine ligand. Owing to the deletion of three amino acids, including the second glycine, the glutamine-binding site cannot form in AtPII, as the first glycine is directly followed by Asp129. With the involvement of three bridging water molecules mimicking the bound ligand, this stunted Q loop can fold independent of glutamine binding, whereas in CrPII the folding of the whole C extension depends on glutamine binding. The three water molecules in AtPII were identified in a reevaluation of the experimental data (see Extended Experimental Procedures).

(C) Sequence alignment of PII C termini of plants and cyanobacteria, highlighting the Q loop consensus motif (bold) and the deletion of three amino acids in the Brassicaceae family. Color labeling shows secondary structure assignment (strands, cyan; helices, pink) according to the CrPII-AtNAGK and AtPII-AtNAGK complex structures. An extended alignment including 55 plant PII sequences is shown in Figure S4.

plants generally requires binding of glutamine to the Q loop, and this was confirmed by examination of *Physcomitrella* and *Oryza* PII proteins. The half-maximal effective concentration of glutamine was determined to be in the millimolar range, which likely reflects the K_d of the glutamine-binding site (EC_{50} of 2.4 mM, 9.2 mM, and 6.6 mM for *Chlamydomonas* PII, *Physcomitrella* PII, and *Oryza* PII, respectively). Notably, these glutamine concentrations are within the range of the estimated concentrations

(2.5–20 mM) of glutamine in different plant species such as tobacco, spinach, and barley (Fritz et al., 2006; Riens et al., 1991; Winter et al., 1993, 1994). Therefore, the low-affinity glutamine-binding mode by the PII Q loop is perfectly tuned for sensing physiologically relevant glutamine levels, which strongly implies regulatory significance. Only when the glutamine concentration in the chloroplast is in the range of several millimolar could PII efficiently activate NAGK and antagonize feedback

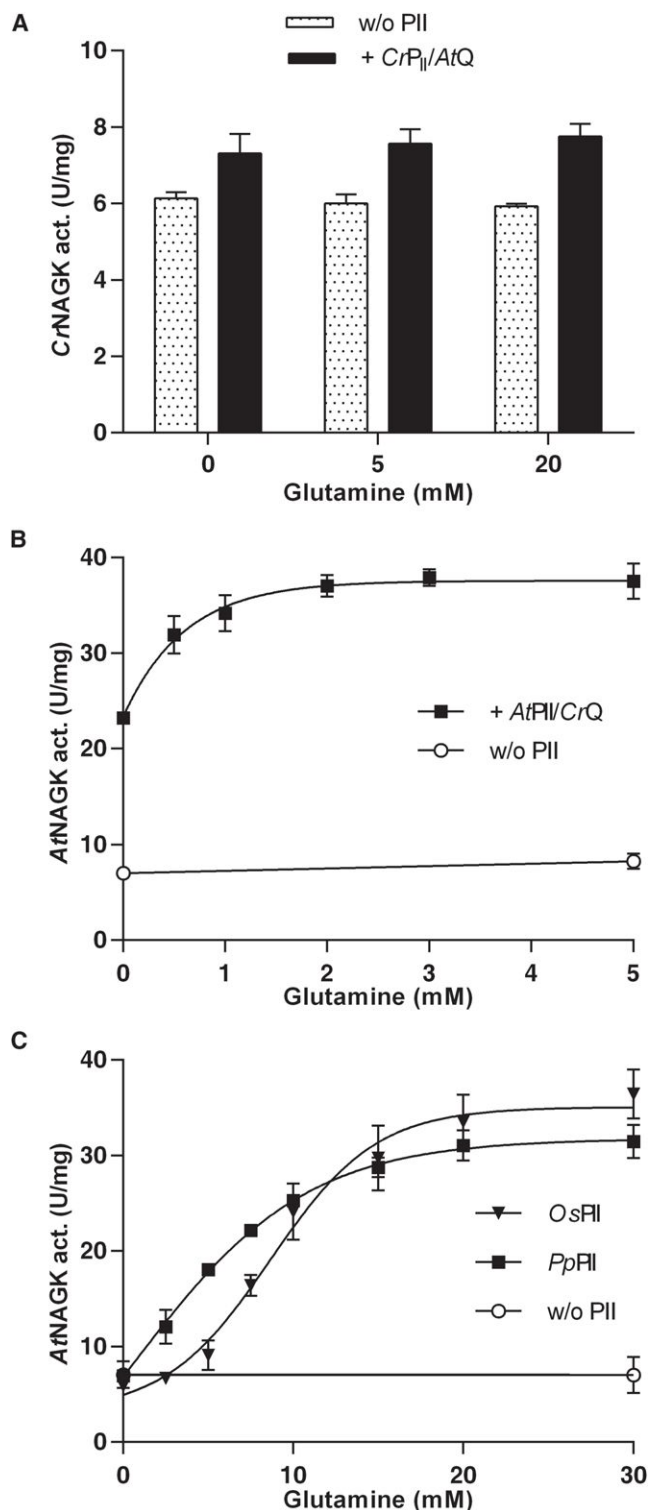


Figure 7. The C-Terminal Segment of P_{II} Proteins from Plants Other Than *A. thaliana* Mediates P_{II} Signaling toward NAGK

(A and B) Chimeric P_{II} proteins consisting of (A) CrP_{II} with the C-terminal Q loop of *A. thaliana* (CrP_{II}/AtQ) and (B) *A. thaliana* P_{II} with the C-terminal Q loop of *C. reinhardtii* (AtP_{II}/CrQ) were tested for glutamine-dependent activation of cognate arginine-feedback-inhibited NAGK.

inhibition by arginine. Indeed, glutamine levels in plants strongly respond to nitrogen status, with amplitude changes of almost 200-fold reported in tobacco leaves (Fritz et al., 2006). In *Chlamydomonas* cells, glutamine levels are 10-fold lower under nitrogen deficiency (Bölling and Fiehn, 2005), accompanied by a similar decline in metabolites of the ornithine/arginine synthesis pathway. Thus, the flow of amino acids into this pathway, which feeds into the arginine pool as well as into polyamines, appears to be controlled by the glutamine level via the P_{II} protein. Arginine has multiple metabolic roles in addition to its roles as a building block in protein biosynthesis, and can also be used as a nitrogen-storage molecule (Llácer et al., 2008). Only at high glutamine levels, which indicate sufficient ammonia supply, is synthesis of nitrogen-storage compounds beneficial.

An intriguing question concerns the evolution of *Brassicaceae* P_{II} proteins. A three amino acid deletion at the Q loop has apparently occurred to generate a glutamine-insensitive variant. Grafting the Q loop from *Chlamydomonas* onto the *A. thaliana* P_{II} body restored glutamine sensitivity at least partially, and shows that *A. thaliana* P_{II} can be converted back to a glutamine-sensing protein. By contrast, the *A. thaliana* C terminus transplanted onto the CrP_{II} body was almost nonfunctional, which indicates that additional modifications in the P_{II} body are required to evolve a functional glutamine-insensitive P_{II} protein. Although glutamine sensing is lost in the *Brassicaceae*, their P_{II} protein is still responsible for the regulation of the ornithine/arginine synthesis pathway; accordingly, P_{II} knockdown lines of *A. thaliana* show impaired accumulation of metabolites from the ornithine/arginine pathway (Ferrario-Méry et al., 2006). The selective pressure that resulted in glutamine-insensitive *Brassicaceae* P_{II} proteins is unknown, but it could point toward a special feature of nitrogen metabolism in this plant family.

Glutamine is the primary nitrogen status reporter in many bacteria (Forchhammer, 2007). There, information about the cellular glutamine status is perceived by the P_{II} signaling system via glutamine-sensitive modification enzymes, such as the uridylyl-transferase/uridylyl-removing enzyme in *E. coli* (Adler et al., 1975; Jiang and Ninfa, 2011). In the development of green algal chloroplasts from cyanobacterial ancestors, evolution has reinvented glutamine sensing by P_{II} signal transducers using a modular concept: a small extension was attached to the C terminus, which controls the interaction of P_{II} with receptors through a subtle network of interprotein interactions for which binding of glutamine is necessary. This is a striking example of convergent

(A) *C. reinhardtii* NAGK with or without the CrP_{II}/AtQ chimera was assayed under standard conditions containing 400 μ M arginine in the absence or presence of 5 or 20 mM glutamine.

(B) Glutamine-dependent activation of arginine-feedback-inhibited AtNAGK by the AtP_{II}/CrQ chimera.

(C) Glutamine-dependent activation of arginine-feedback-inhibited AtNAGK by *Physcomitrella* P_{II} (PpP_{II}) or *Oryza* P_{II} (OsP_{II}).

Assays in (B) and (C) contained the NAGK feedback inhibitor arginine at 20 mM and increasing concentrations of glutamine, as indicated. The activity of NAGK in the absence of any P_{II} protein (without P_{II}) at 0 and 5 mM glutamine (B) and 0 and 30 mM glutamine (C) is indicated.

The standard deviation from triplicate measurements is indicated by error bars.

evolution in one of nature's most abundant signaling proteins. By the identification of P_{II} as a glutamine sensor, this work opens the way to a deeper understanding of nitrogen homeostasis in plants and may help optimize plant breeding for crop yield improvement.

EXPERIMENTAL PROCEDURES

Detailed methods are given in Extended Experimental Procedures.

Cloning, Expression, and Purification of P_{II} Proteins

Recombinant P_{II} from *A. thaliana* (AtP_{II}) and from *C. reinhardtii* P_{II} (CrP_{II}) with a C-terminally fused Strep-tag II sequence were overexpressed and purified as described previously (Beez et al., 2009; Ermilova et al., 2013). The CrP_{II}-ΔC-coding gene was constructed using standard PCR techniques. AtP_{II}/CrQ chimera, *P. patens* PpP_{II}, and *O. sativa* OsP_{II} genes were commercially synthesized. All genes were cloned into the Strep-tag fusion vector pASK-IBA3plus. Overexpression and purification of the P_{II} proteins by affinity chromatography on Strep-Tactin columns were described previously (Heinrich et al., 2004). Biochemical properties of the purified CrP_{II} and CrP_{II}-ΔC proteins are shown in Figure S3.

Cloning, Expression, and Purification of NAGK Proteins

Recombinant AtNAGK was overexpressed and purified as described previously (Beez et al., 2009). For recombinant CrNAGK, a gene corresponding to mature chloroplast-localized CrNAGK was commercially synthesized, cloned into vector pET15b, and purified as recombinant AtNAGK. Analysis of purified CrNAGK and AtNAGK by gel filtration is shown in Figure S2. A multiple sequence alignment of plant and cyanobacterial NAGK proteins, including predicted chloroplast transit peptides for CrNAGK and AtNAGK, is shown in Figure S1.

Surface Plasmon Resonance Analysis

SPR experiments were performed using a Biacore X biosensor system at 25°C in HEPES-buffered saline (HBS buffer) as described previously (Maheswaran et al., 2004). His₆-NAGK was immobilized on flow cell (FC) 2 to a density of approximately 3,000 resonance units (RUs). The analyte solutions contained P_{II} proteins (100 nM) in HBS buffer and were incubated with various effector molecules on ice for 5 min and then injected (50 μl) into FC1 (control) and FC2 of the sensor chip. The specific binding of P_{II} to NAGK was recorded as the response signal difference FC2 – FC1 (ΔRU).

Isothermal Titration Calorimetry

ITC experiments were performed on a MicroCal VP-ITC instrument in 10 mM potassium phosphate (pH 7.5), 100 mM NaCl, and 2 mM MgCl₂ at 20°C.

Enzymatic Assay for NAGK Activity

A coupled enzyme assay was used to determine NAGK activity in which the production of ADP was coupled to the oxidation of NADH (Beez et al., 2009). The reaction mixture contained 2.4 μg of P_{II} protein; the reaction was started by adding 3 μg of NAGK. Means of triplicate experimental determinations are shown. From the velocity slopes, the catalytic constants were calculated using the GraphPad Prism 6.01 software program.

Crystallization, Sample Preparation, Data Collection, and Structure Determination

Crystallization trials were performed in a standard vapor-diffusion setup. Crystals of the CrP_{II} apo form grew in a solution containing 3 mg/ml CrP_{II} in 10 mM Tris (pH 7.5), 100 mM NaCl, 2 mM ATP, 2 mM MgCl₂, 5 mM Gln, and 10% (v/v) glycerol and a reservoir solution containing 0.15 M (NH₄)₂SO₄, 0.1 M HEPES (pH 7.0), and 20% (w/v) PEG 4000. For the preparation of the CrP_{II}•Mg-ATP•2-OG complex, these crystals were soaked in a droplet of reservoir solution supplemented with 10 mM ATP, 10 mM MgCl₂, 10 mM 2-OG, and 20% (v/v) glycerol. Crystals of the CrP_{II}-AtNAGK complex grew in a solution con-

taining 2 mg/ml protein (with a 1:2 molar ratio of CrP_{II}:AtNAGK) in 10 mM Tris (pH 7.8), 100 mM NaCl, 40 mM Arg, 10 mM Gln, 2 mM MgCl₂, 2 mM ADP, 10 mM NAG, and 5% (v/v) glycerol and a reservoir solution containing 200 mM NaCl, 100 mM Na/K phosphate (pH 6.2), and 50% (v/v) PEG 200. Data were collected under cryogenic conditions.

The CrP_{II} structure was solved by molecular replacement using the SeP_{II} structure 2XZW. The CrP_{II}-AtNAGK complex structure was solved on the basis of the AtP_{II}-AtNAGK structure 2RD5. Data processing and refinement statistics are summarized in Table S1.

ACCESSION NUMBERS

The Protein Data Bank ID codes for the structures reported herein are 4USH (CrP_{II} apo), 4USI (CrP_{II} with ligands), and 4USJ (CrP_{II}-AtNAGK complex).

SUPPLEMENTAL INFORMATION

Supplemental Information includes Extended Experimental Procedures, four figures, and one table and can be found with this article online at <http://dx.doi.org/10.1016/j.cell.2014.10.015>.

AUTHOR CONTRIBUTIONS

V.-R.C. biochemically analyzed CrP_{II}/CrNAGK, carried out the crystallization and diffraction experiments, and solved and analyzed the crystal structures; E.E. evaluated results of the CrP_{II}/AtNAGK experiments; T.L. carried out the CrP_{II}/AtNAGK experiments; J.L. designed and performed the cloning experiments; E.M. prepared and tested the OsP_{II} and PpP_{II} proteins; C.H. purified the proteins and performed enzyme assays; M.D.H. analyzed sequences, solved and analyzed the crystal structures, and wrote the manuscript; and K.F. designed, analyzed, and evaluated the experiments and wrote the manuscript.

ACKNOWLEDGMENTS

This work was funded by the Deutsche Forschungsgemeinschaft (grants Fo195/9-1 and Fo195/10/11 to K.F.) and by a grant of the Russian Foundation for Basic Research (N 13-04-00087 to E.E.). V.-R.C. and M.D.H. are grateful to Andrei Lupas for continuous support. We thank Karen Brune for careful linguistic editing of the paper and Oleksandra Fokina for helpful experimental advice, acknowledge the skillful assistance of Reinhard Albrecht and Kerstin Bär in crystallographic sample preparation, and are grateful for the excellent support by the staff of Swiss Light Source beamline X10SA.

Received: May 23, 2014

Revised: August 13, 2014

Accepted: September 26, 2014

Published: November 20, 2014

REFERENCES

- Adler, S.P., Purich, D., and Stadtman, E.R. (1975). Cascade control of *Escherichia coli* glutamine synthetase. Properties of the P_{II} regulatory protein and the uridylyltransferase-uridylyl-removing enzyme. *J. Biol. Chem.* 250, 6264–6272.
- Beez, S., Fokina, O., Herrmann, C., and Forchhammer, K. (2009). *N*-acetyl-L-glutamate kinase (NAGK) from oxygenic phototrophs: P_{II} signal transduction across domains of life reveals novel insights in NAGK control. *J. Mol. Biol.* 389, 748–758.
- Bölling, C., and Fiehn, O. (2005). Metabolite profiling of *Chlamydomonas reinhardtii* under nutrient deprivation. *Plant Physiol.* 139, 1995–2005.
- Chellamuthu, V.R., Alva, V., and Forchhammer, K. (2013). From cyanobacteria to plants: conservation of P_{II} functions during plastid evolution. *Planta* 237, 451–462.

- Ermilova, E., Lapina, T., Zalutskaya, Z., Minaeva, E., Fokina, O., and Forchhammer, K. (2013). P_{II} signal transduction protein in *Chlamydomonas reinhardtii*: localization and expression pattern. *Protist* 164, 49–59.
- Ferrario-Méry, S., Bouvet, M., Leleu, O., Savino, G., Hodges, M., and Meyer, C. (2005). Physiological characterisation of *Arabidopsis* mutants affected in the expression of the putative regulatory protein P_{II}. *Planta* 223, 28–39.
- Ferrario-Méry, S., Besin, E., Pichon, O., Meyer, C., and Hodges, M. (2006). The regulatory P_{II} protein controls arginine biosynthesis in *Arabidopsis*. *FEBS Lett.* 580, 2015–2020.
- Ferrario-Méry, S., Meyer, C., and Hodges, M. (2008). Chloroplast nitrite uptake is enhanced in *Arabidopsis* P_{II} mutants. *FEBS Lett.* 582, 1061–1066.
- Fokina, O., Chellamuthu, V.R., Forchhammer, K., and Zeth, K. (2010). Mechanism of 2-oxoglutarate signaling by the *Synechococcus elongatus* P_{II} signal transduction protein. *Proc. Natl. Acad. Sci. USA* 107, 19760–19765.
- Forchhammer, K. (2007). Glutamine signalling in bacteria. *Front. Biosci.* 12, 358–370.
- Forchhammer, K. (2008). P_{II} signal transducers: novel functional and structural insights. *Trends Microbiol.* 16, 65–72.
- Forchhammer, K., and Tandeau de Marsac, N. (1994). The P_{II} protein in the cyanobacterium *Synechococcus* sp. strain PCC 7942 is modified by serine phosphorylation and signals the cellular N-status. *J. Bacteriol.* 176, 84–91.
- Foyer, C.H., Parry, M., and Noctor, G. (2003). Markers and signals associated with nitrogen assimilation in higher plants. *J. Exp. Bot.* 54, 585–593.
- Fritz, C., Mueller, C., Matt, P., Feil, R., and Stitt, M. (2006). Impact of the C-N status on the amino acid profile in tobacco source leaves. *Plant Cell Environ.* 29, 2055–2076.
- Harris, E.H. (2001). *Chlamydomonas* as a model organism. *Annu. Rev. Plant Physiol. Plant Mol. Biol.* 52, 363–406.
- Heinrich, A., Maheswaran, M., Ruppert, U., and Forchhammer, K. (2004). The *Synechococcus elongatus* P signal transduction protein controls arginine synthesis by complex formation with N-acetyl-L-glutamate kinase. *Mol. Microbiol.* 52, 1303–1314.
- Heinrich, A., Woyda, K., Brauburger, K., Meiss, G., Detsch, C., Stülke, J., and Forchhammer, K. (2006). Interaction of the membrane-bound GlnK-AmtB complex with the master regulator of nitrogen metabolism TnrA in *Bacillus subtilis*. *J. Biol. Chem.* 281, 34909–34917.
- Hesketh, A., Fink, D., Gust, B., Rexer, H.U., Scheel, B., Chater, K., Wohlleben, W., and Engels, A. (2002). The GlnD and GlnK homologues of *Streptomyces coelicolor* A3(2) are functionally dissimilar to their nitrogen regulatory system counterparts from enteric bacteria. *Mol. Microbiol.* 46, 319–330.
- Hsieh, M.H., Lam, H.M., van de Loo, F.J., and Coruzzi, G. (1998). A P_{II}-like protein in *Arabidopsis*: putative role in nitrogen sensing. *Proc. Natl. Acad. Sci. USA* 95, 13965–13970.
- Huergo, L.F., Chandra, G., and Merrick, M. (2013). P_{II} signal transduction proteins: nitrogen regulation and beyond. *FEMS Microbiol. Rev.* 37, 251–283.
- Jiang, P., and Ninfa, A.J. (2009a). α -ketoglutarate controls the ability of the *Escherichia coli* P_{II} signal transduction protein to regulate the activities of NRII (NtrB) but does not control the binding of P_{II} to NRII. *Biochemistry* 48, 11514–11521.
- Jiang, P., and Ninfa, A.J. (2009b). Sensation and signaling of α -ketoglutarate and adenylate energy charge by the *Escherichia coli* P_{II} signal transduction protein require cooperation of the three ligand-binding sites within the P_{II} trimer. *Biochemistry* 48, 11522–11531.
- Jiang, P., and Ninfa, A.J. (2011). A source of ultrasensitivity in the glutamine response of the bicyclic cascade system controlling glutamine synthetase adenylation state and activity in *Escherichia coli*. *Biochemistry* 50, 10929–10940.
- Jiang, P., Zucker, P., Atkinson, M.R., Kamberov, E.S., Tirasophon, W., Chandran, P., Scheffke, B.R., and Ninfa, A.J. (1997). Structure/function analysis of the P_{II} signal transduction protein of *Escherichia coli*: genetic separation of interactions with protein receptors. *J. Bacteriol.* 179, 4342–4353.
- Keys, A.J. (2006). The re-assimilation of ammonia produced by photorespiration and the nitrogen economy of C3 higher plants. *Photosynth. Res.* 87, 165–175.
- Lea, P.J., and Mifflin, B.J. (2010). Nitrogen assimilation and its relevance to crop improvement. In *Annual Plant Reviews, Volume 42: Nitrogen Metabolism in Plants in the Post-Genomic Era*, Foyer C.H. and Zhang H., eds. (Oxford: Wiley-Blackwell), pp. 1–40.
- Leigh, J.A., and Dodsworth, J.A. (2007). Nitrogen regulation in bacteria and archaea. *Annu. Rev. Microbiol.* 61, 349–377.
- Liácer, J.L., Contreras, A., Forchhammer, K., Marco-Marín, C., Gil-Ortiz, F., Maldonado, R., Fita, I., and Rubio, V. (2007). The crystal structure of the complex of P_{II} and acetylglutamate kinase reveals how P_{II} controls the storage of nitrogen as arginine. *Proc. Natl. Acad. Sci. USA* 104, 17644–17649.
- Liácer, J.L., Fita, I., and Rubio, V. (2008). Arginine and nitrogen storage. *Curr. Opin. Struct. Biol.* 18, 673–681.
- Maheswaran, M., Urbanke, C., and Forchhammer, K. (2004). Complex formation and catalytic activation by the P_{II} signaling protein of N-acetyl-L-glutamate kinase from *Synechococcus elongatus* strain PCC 7942. *J. Biol. Chem.* 279, 55202–55210.
- Maier, S., Schleberger, P., Lü, W., Wacker, T., Pflüger, T., Litz, C., and Andrade, S.L.A. (2011). Mechanism of disruption of the Amt-GlnK complex by P_{II}-mediated sensing of 2-oxoglutarate. *PLoS ONE* 6, e26327.
- Mizuno, Y., Berenger, B., Moorhead, G.B., and Ng, K.K. (2007a). Crystal structure of *Arabidopsis* P_{II} reveals novel structural elements unique to plants. *Biochemistry* 46, 1477–1483.
- Mizuno, Y., Moorhead, G.B., and Ng, K.K. (2007b). Structural basis for the regulation of N-acetylglutamate kinase by P_{II} in *Arabidopsis thaliana*. *J. Biol. Chem.* 282, 35733–35740.
- Ninfa, A.J., and Atkinson, M.R. (2000). P_{II} signal transduction proteins. *Trends Microbiol.* 8, 172–179.
- Ninfa, A.J., and Jiang, P. (2005). P_{II} signal transduction proteins: sensors of α -ketoglutarate that regulate nitrogen metabolism. *Curr. Opin. Microbiol.* 8, 168–173.
- Nunes-Nesi, A., Fernie, A.R., and Stitt, M. (2010). Metabolic and signaling aspects underpinning the regulation of plant carbon nitrogen interactions. *Plant* 3, 973–996.
- Radchenko, M., and Merrick, M. (2011). The role of effector molecules in signal transduction by P_{II} proteins. *Biochem. Soc. Trans.* 39, 189–194.
- Radchenko, M.V., Thornton, J., and Merrick, M. (2013). P_{II} signal transduction proteins are ATPases whose activity is regulated by 2-oxoglutarate. *Proc. Natl. Acad. Sci. USA* 110, 12948–12953.
- Riens, B., Lohaus, G., Heineke, D., and Heldt, H.W. (1991). Amino acid and sucrose content determined in the cytosolic, chloroplastic, and vacuolar compartments and in the phloem sap of spinach leaves. *Plant Physiol.* 97, 227–233.
- Sant'Anna, F.H., Trentini, D.B., de Souto Weber, S., Cecagno, R., da Silva, S.C., and Schrank, I.S. (2009). The P_{II} superfamily revised: a novel group and evolutionary insights. *J. Mol. Evol.* 68, 322–336.
- Smith, C.S., Weljie, A.M., and Moorhead, G.B. (2003). Molecular properties of the putative nitrogen sensor P_{II} from *Arabidopsis thaliana*. *Plant J.* 33, 353–360.
- Smith, C.S., Morrice, N.A., and Moorhead, G.B. (2004). Lack of evidence for phosphorylation of *Arabidopsis thaliana* P_{II}: implications for plastid carbon and nitrogen signaling. *Biochim. Biophys. Acta* 1699, 145–154.
- Sugiyama, K., Hayakawa, T., Kudo, T., Ito, T., and Yamaya, T. (2004). Interaction of N-acetylglutamate kinase with a P_{II}-like protein in rice. *Plant Cell Physiol.* 45, 1768–1778.

- Truan, D., Huergo, L.F., Chubatsu, L.S., Merrick, M., Li, X.D., and Winkler, F.K. (2010). A new P_{II} protein structure identifies the 2-oxoglutarate binding site. *J. Mol. Biol.* 400, 531–539.
- Uhrig, R.G., Ng, K.K., and Moorhead, G.B. (2009). PII in higher plants: a modern role for an ancient protein. *Trends Plant Sci.* 14, 505–511.
- Winter, H., Robinson, D.G., and Heldt, H.W. (1993). Subcellular volumes and metabolite concentrations in barley leaves. *Planta* 191, 180–190.
- Winter, H., Robinson, D.G., and Heldt, H.W. (1994). Subcellular volumes and metabolite concentrations in spinach leaves. *Planta* 193, 530–535.
- Zeth, K., Fokina, O., and Forchhammer, K. (2014). Structural basis and target-specific modulation of ADP sensing by the *Synechococcus elongatus* P_{II} signaling protein. *J. Biol. Chem.* 289, 8960–8972.
- Zheng, Z.L. (2009). Carbon and nitrogen nutrient balance signaling in plants. *Plant Signal. Behav.* 4, 584–591.

An Integrated Approach Reveals Regulatory Controls on Bacterial Translation Elongation

Arvind R. Subramaniam,^{1,2,*} Brian M. Zid,^{1,2} and Erin K. O'Shea^{1,2,3,4,*}

¹Faculty of Arts and Sciences Center for Systems Biology

²Department of Molecular and Cellular Biology

³Department of Chemistry and Chemical Biology

⁴Howard Hughes Medical Institute

Harvard University, Cambridge, MA 02138, USA

*Correspondence: asubram@fas.harvard.edu (A.R.S.), erin_oshea@harvard.edu (E.K.O.)

<http://dx.doi.org/10.1016/j.cell.2014.10.043>

SUMMARY

Ribosomes elongate at a nonuniform rate during translation. Theoretical models and experiments disagree on the in vivo determinants of elongation rate and the mechanism by which elongation rate affects protein levels. To resolve this conflict, we measured transcriptome-wide ribosome occupancy under multiple conditions and used it to formulate a whole-cell model of translation in *E. coli*. Our model predicts that elongation rates at most codons during nutrient-rich growth are not limited by the intracellular concentrations of aminoacyl-tRNAs. However, elongation pausing during starvation for single amino acids is highly sensitive to the kinetics of tRNA aminoacylation. We further show that translation abortion upon pausing accounts for the observed ribosome occupancy along mRNAs during starvation. Abortion reduces global protein synthesis, but it enhances the translation of a subset of mRNAs. These results suggest a regulatory role for aminoacylation and abortion during stress, and our study provides an experimentally constrained framework for modeling translation.

INTRODUCTION

Protein synthesis begins with initiation by ribosomes on an mRNA and is followed by a sequence of elongation steps during which amino acids are added to the growing polypeptide chain. Initiation is the rate-limiting step for the translation of most mRNAs during nutrient-rich growth (Jacques and Dreyfus, 1990). However, stressful perturbations such as amino acid starvation or transgene overexpression can decrease the elongation rate of ribosomes and affect protein levels (Subramaniam et al., 2013a; Varenne et al., 1984; Welch et al., 2009). Biophysical modeling of translation can be used to infer the quantitative effect of these stressful perturbations on initiation rate, elongation rate, and the expression level of proteins (Shah et al., 2013). More generally, modeling integrates our biochemical knowledge of translation and thus enables identifica-

tion of novel regulatory processes when incorporation of known mechanisms is insufficient to recapitulate experimental measurements.

Biophysical modeling of translation has been greatly aided by the development of ribosome profiling, which involves deep sequencing of ribosome-protected mRNA fragments to measure ribosome occupancy on mRNAs (Ingolia et al., 2009). Ribosome profiling has also produced surprising results that challenge two central assumptions in current theoretical models of translation. First, theoretical models assume that elongation rates at codons are directly proportional to the intracellular concentration of cognate aminoacyl-tRNAs (aa-tRNAs) during nutrient-rich growth (Shah et al., 2013; Tuller et al., 2010). However, this assumption is not supported by ribosome profiling—although concentrations of different tRNAs vary over greater than a 10-fold range in bacteria and yeast (Dong et al., 1996; Tuller et al., 2010), the measured ribosome occupancy at codons varies less than 2-fold in these organisms during nutrient-rich growth (Li et al., 2012; Qian et al., 2012). Differential aminoacylation of tRNAs is also unlikely to underlie this discrepancy, as most tRNA species are >70% aminoacylated during nutrient-rich growth (Yegian et al., 1966). Second, most models assume that a decrease in ribosome elongation rate on an mRNA affects the level of the corresponding protein by causing a traffic jam of trailing ribosomes (Shah et al., 2013; Zhang et al., 1994). However, except for a few atypical mRNAs (Guydosh and Green, 2014; Li et al., 2012), traffic jams at ribosome pause sites have not been observed in vivo, and the effect of ribosome traffic jams on protein level remains unclear. Together, these results suggest that current theoretical models do not include the full set of mechanistic ingredients that is necessary for accurate modeling of in vivo translation, especially as it pertains to the elongation stage.

To identify mechanistic features of the elongation stage that enable accurate biophysical modeling of translation, we combined transcriptome-scale and reporter-based experiments with whole-cell computational modeling. This approach enabled us to decipher the contribution of various molecular processes to the elongation rate of ribosomes and the synthesis rate of proteins. We found that the differences in the intracellular concentration of tRNAs and the occurrence of ribosome traffic jams at pause sites, both of which have been key ingredients in previous theoretical models (Shah et al., 2013; Tuller

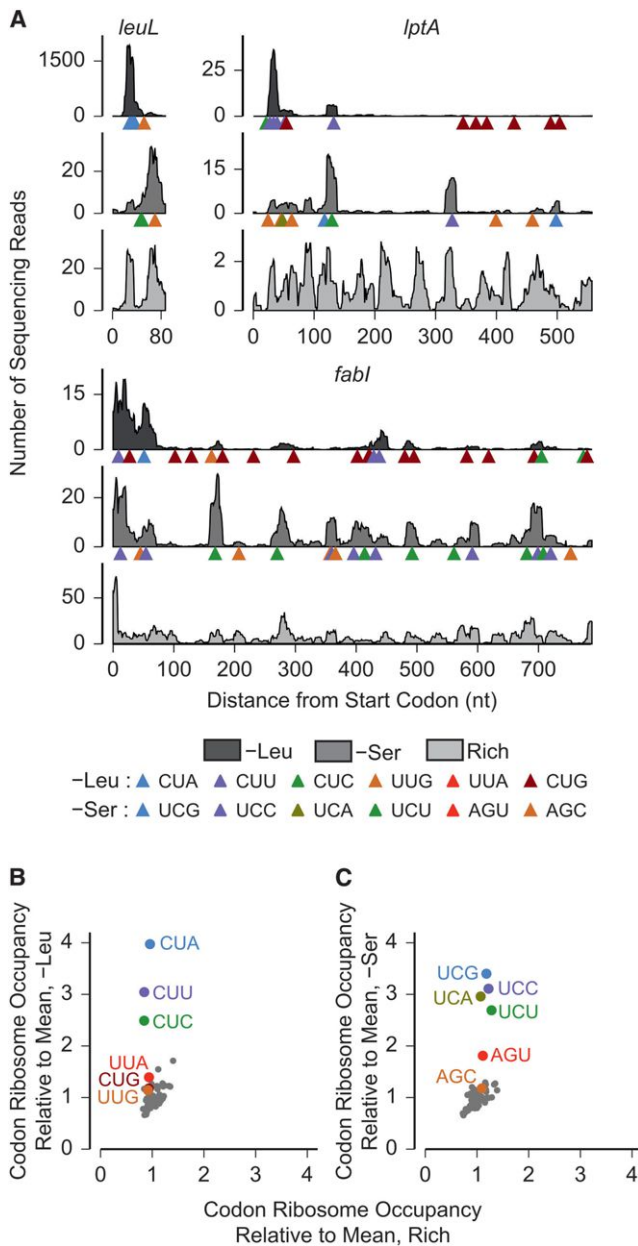


Figure 1. Changes in Ribosome Occupancy upon Starvation for Single Amino Acids

(A) Measured ribosome occupancy along three *E. coli* genes: *leuL*, *lptA*, and *fabI* during leucine starvation (-Leu), serine starvation (-Ser), and amino-acid-rich growth (Rich). The horizontal axis extends from the start codon to the stop codon for each gene. Triangles indicate the positions of leucine (serine) codons along the coding sequence in the leucine (serine) starvation case.

(B and C) Measured ribosome occupancy at the 61 sense codons averaged across the transcriptome. Start and stop codons are not shown. Standard errors of mean are smaller than data markers.

See also Figure S1.

et al., 2010), are insufficient to predict the measured ribosome occupancy in our experiments. We find that the molecular process of aminoacylation has a critical role in protein synthesis

through modulation of ribosome elongation rates during stress. Further, we find that premature termination of translation before the synthesis of full-length proteins (henceforth referred to as translation abortion) determines both the ribosome occupancy along mRNAs and protein expression during stress. More generally, our work illustrates the usefulness of integrating deep-sequencing experimental methods such as ribosome profiling together with quantitative whole-cell modeling to reveal the experimentally relevant regimes of the a priori large parameter space in biophysical models of complex cellular processes.

RESULTS

Changes in Ribosome Occupancy upon Starvation for Single Amino Acids

We previously found that, during starvation for single amino acids in *E. coli*, the presence of certain codons cognate to the limiting amino acid can decrease the protein synthesis rate by up to 100-fold (Subramaniam et al., 2013a). Starvation for single amino acids decreases the concentration of the cognate aminoacyl (aa) tRNAs (Dittmar et al., 2005; Sørensen, 2001; Sørensen et al., 2005), thus decreasing the elongation rate of ribosomes at cognate codons. Based on these observations, we used single amino acid starvation as an experimental condition to develop constraints for biophysical models of protein synthesis in an elongation-limited regime of translation.

To characterize the effect of amino acid starvation on translation, we performed ribosome profiling in *E. coli* after 30 min of starvation for each of two amino acids: leucine and serine. For comparison with an initiation-limited regime of translation, we also performed ribosome profiling on cells grown in rich-defined medium with all 20 amino acids. The distribution of ribosome footprints along mRNAs was highly variable during nutrient-rich growth (Figure 1A), consistent with earlier observations (Oh et al., 2011). Starvation for either leucine or serine caused a pronounced change in the distribution of ribosome footprints along individual mRNAs (Figure 1A). Consistent with previous observations (Li et al., 2012), the transcriptome-averaged ribosome occupancy differed less than 2-fold across the 61 codons during nutrient-rich growth (Figures 1B and 1C, horizontal axis) and did not systematically vary with tRNA abundance (Figure S1A available online). Upon leucine or serine starvation, the average ribosome occupancy increased at leucine and serine codons, respectively, but this increase was not uniform (Figures 1B and 1C, vertical axis). The three leucine codons CUA, CUC, and CUU had 2.5- to 4-fold higher ribosome occupancy than the average during leucine starvation, whereas the ribosome occupancy at the remaining three leucine codons CUG, UUA, and UUG was comparable to that of noncognate codons. Similarly, ribosome occupancy at the four serine codons UCA, UCC, UCG, and UCU was markedly higher than at the two serine codons AGC and AGU during serine starvation. We attributed the increase in ribosome occupancy at leucine and serine codons to a decrease in the ribosome elongation rate, as we did not observe an increase in total mRNA density at these codons upon leucine or serine starvation (Figures S1B and S1C).

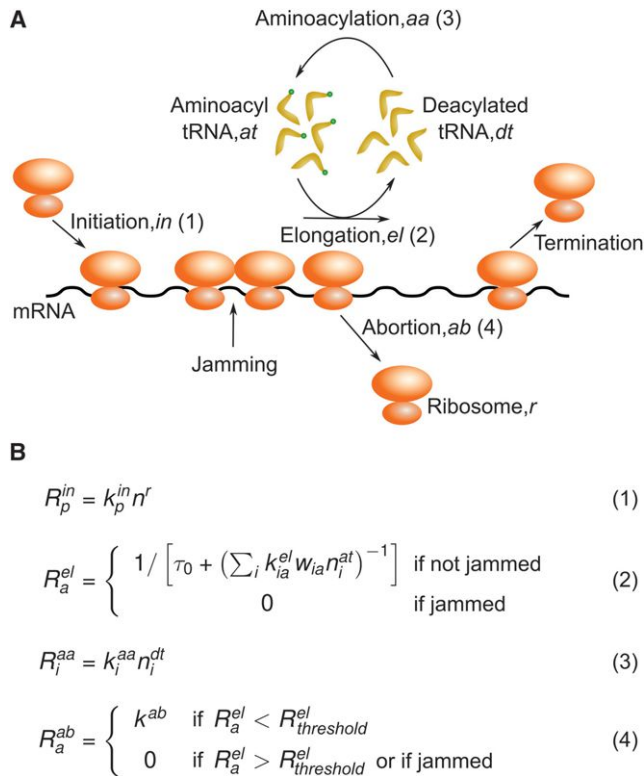


Figure 2. A Transcriptome-Scale Biophysical Model of Translation
(A) Schematic of the four cellular processes modeled (initiation, elongation, aminoacylation, abortion) and the molecular species considered in the biophysical model.

(B) Reaction rates, R_i , for the cellular processes shown in (A). The superscripts following R_i refer to the abbreviations for the cellular processes in (A). Subscript indices are used for distinct molecular species of the same kind (mRNA – p , tRNA – i , codon – a). The intracellular concentrations of molecular species and the values for rate constants in our whole-cell simulation are in Table S1.

See also Data S1.

A Transcriptome-Scale Biophysical Model of Translation in *E. coli*

To systematically evaluate the consistency of different biophysical models with our measurements of ribosome occupancy, we formulated a generalized kinetic model of protein synthesis in *E. coli* that accounted for four different molecular processes that we found to play a critical role in determining ribosome occupancy and protein expression: initiation, elongation, aminoacylation, and abortion (Figure 2A). Other molecular processes that are not expected to limit protein synthesis rate under the conditions of our study, such as termination at stop codons and ribosome recycling, were assumed to be instantaneous (rates set to infinity) for the sake of simplicity.

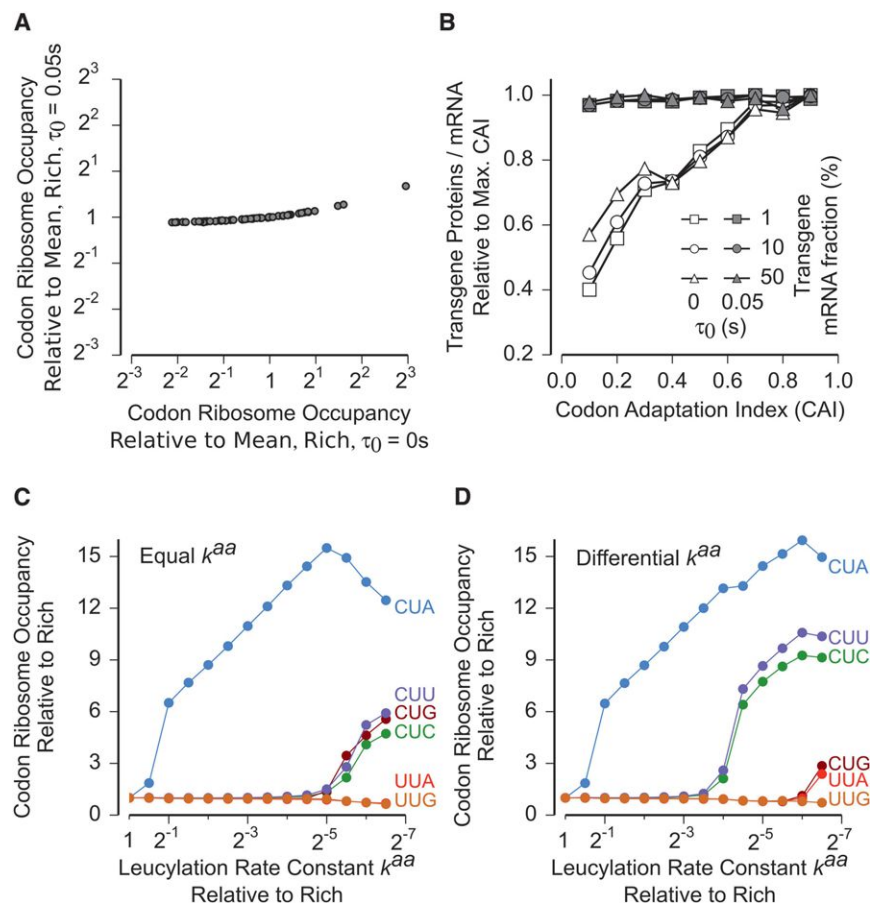
The core features of our biophysical model of protein synthesis are summarized by the kinetic rate equations for the four molecular processes (Figure 2B). The full list of parameter values used in our biophysical model is given in Table S1. We implemented our model as a continuous-time, discrete-state process at the

whole-cell level for *E. coli* by adapting the source code from a recent computational study in yeast (Shah et al., 2013). During balanced nutrient-rich growth, the availability of free ribosomes for initiation is rate limiting for protein synthesis (Vind et al., 1993). We modeled this empirical observation by taking the initiation rate R^{in} to be proportional to n^r , the number of free ribosomes in the cell (Equation 1 in Figure 2B). The initiation rate constant k_p^{in} for each mRNA species p was estimated from our ribosome profiling measurements and the macromolecular composition and synthesis rates that have been measured with high accuracy in *E. coli* (Bremer and Dennis, 2008) (Extended Experimental Procedures).

Intraribosomal Events Limit the Rate of Elongation during Nutrient-Rich Growth

The elongation cycle is a multistep process that begins with the delivery of the aa-tRNA (in ternary complex with Ef-Tu and GTP) to the ribosome. This step is followed by a series of intraribosomal events that include kinetic proofreading, peptidyl transfer, and translocation of the ribosome to the next codon (Wintermeyer et al., 2004). We considered a minimal model of elongation (Elf et al., 2003) that is composed of two effective steps (Equation 2 in Figure 2B). The first step accounts for the arrival of cognate aa-tRNAs at the ribosome, and its rate is proportional to the intracellular concentration of aa-tRNA, n^{at} . The proportionality factor k^{el} is the second-order rate constant (k_{cat}/K_M) for association between the ribosome and the aa-tRNA. The second step accounts for all intraribosomal events that follow the arrival of the cognate aa-tRNA at the ribosome, and the cumulative rate of these events does not depend on the intracellular concentration of aa-tRNA. For simplicity, we set the total rate of this second step to a uniform value $1/\tau_0$ for all codons. Finally, when translocation is prevented by the presence of a leading ribosome (henceforth referred to as jamming), the elongation rate is set to zero in our model.

If τ_0 (the rate constant describing all intraribosomal events following arrival of the aa-tRNA) has the same value for all codons, the experimental observation that ribosome occupancies at codons do not vary inversely with the corresponding tRNA concentrations during nutrient-rich growth (Figure S1A) now implies a straightforward mechanistic constraint in our model—that the rate of intraribosomal events limits the overall rate of ribosome elongation under these conditions, i.e., $\tau_0 \gg 1/k^{el}n^{at}$ in Equation 2 of Figure 2B. Together, with the measured average elongation rate of ribosomes of ~ 20 codons per second (Bremer and Dennis, 2008), we can then infer that $\tau_0 \approx 0.05$ s. In the more general case in which τ_0 is not uniform for all codons, most of the codons still need to have $\tau_0 \gg 1/k^{el}n^{at}$ to recapitulate the observed lack of (inverse) proportionality between ribosome occupancy and tRNA concentration. Our conclusion that $\tau_0 \gg 1/k^{el}n^{at}$, based solely on ribosome occupancy measurements, is also consistent with the typical in vitro value of $k^{el} = 2 \times 10^7 \text{ M}^{-1}\text{s}^{-1}$ (Bilgin and Ehrenberg, 1994; Pavlov and Ehrenberg, 1996) and the in vivo concentrations of tRNAs (n^{at}) during nutrient-rich growth (Bremer and Dennis, 2008; Dong et al., 1996) (Data S1). Using these values of k^{el} and n^{at} , the median value of $1/k^{el}n^{at}$ for the 61 sense codons is 0.0036 s, and all codons had $1/k^{el}n^{at} < 0.013$ s ($\ll \tau_0 \approx 0.05$ s).



To test the consistency of our whole-cell model with ribosome profiling measurements, we simulated our model with the above-constrained rate constants for elongation. With $\tau_0 = 0.05$ s and $k^{el} = 2 \times 10^7$ M⁻¹s⁻¹, we observed less than 2-fold variation (SD/mean = 10%) in ribosome occupancy across the 61 codons (Figure 3A, vertical axis). By contrast, when we set the timescale of intraribosomal events to be zero ($\tau_0 = 0$ s), as assumed in previous work (Shah et al., 2013; Tuller et al., 2010), ribosome occupancy varied 34-fold (SD/mean = 105%) across the 61 codons (Figure 3A, horizontal axis), even though for both values of τ_0 (0.05 and 0 s), the average elongation rate (R^{el}) was ~ 20 codons s⁻¹. Thus, we conclude that ribosome occupancy at codons during nutrient-rich growth of *E. coli*, as measured using ribosome profiling, is consistent with a model in which intraribosomal events, rather than the arrival of aa-tRNA to the ribosome A site, are the slowest steps in the elongation cycle for most codons.

Highly expressed genes in *E. coli*, such as those coding for the translation machinery, display a characteristic codon bias (called the major codon bias) toward codons that are decoded by tRNA isoacceptors with high intracellular concentration (Andersson and Kurland, 1990). Motivated by this observation, biophysical models often predict that optimizing the codon usage of an overexpressed transgene by enriching for codons decoded by abundant tRNA isoacceptors can significantly improve the yield of its protein (Shah et al., 2013; Zhang et al., 1994). Because this question is of

Figure 3. Elongation and Aminoacylation Kinetics Determine Ribosome Occupancy at Codons

(A) Mean ribosome occupancy at the 61 sense codons averaged across the transcriptome during nutrient-rich growth calculated from whole-cell model. Simulations were run with the time for intraribosomal events at a single codon, τ_0 , set to either 0 s (horizontal) or 0.05 s (vertical). The value of k^{el} was chosen such that the mean elongation rate of ribosomes R^{el} was approximately equal to the experimentally measured value of 20 s⁻¹ in both cases. (B) Amount of transgene proteins produced per mRNA upon overexpression during nutrient-rich growth calculated from whole-cell model as a function of codon adaptation index (CAI) and the transgene fraction. All data points corresponding to a single transgene mRNA fraction were normalized by the data point at CAI = 0.9. (C and D) Mean ribosome occupancy at the six leucine codons as a function of leucylation rate constant calculated from whole-cell model. The leucylation rate constants of the five leucine tRNA isoacceptors were set either equal (C) or different (D). In the differential case (D), the leucylation rate constants were in the proportion 1.5: 0.5: 1: 0.5: 0.5 (Leu1 through Leu5). See also Figure S2.

significant interest in biotechnological applications (Gustafsson et al., 2004), we re-examined it in the light of our conclusion that intraribosomal events, rather than the arrival of aa-tRNA, limit the rate of elongation during nutrient-rich growth.

We simulated the expression of three transgenes at different fractions of the cellular transcriptome and with various codon bias (Extended Experimental Procedures and Data S1). When we assumed that the elongation rate at codons is proportional to the concentration of aa-tRNAs (by setting $\tau_0 = 0$ s in Equation 2 of Figure 2B), we found that increasing codon bias can improve protein expression from the transgenes by >2-fold (Figure 3B, open markers). However, this effect of codon bias was essentially absent when elongation rates at codons were limited by intraribosomal events (Figure 3B, filled markers). We observed a similar effect of intraribosomal events when we simulated a model for yeast translation (Shah et al., 2013) (Figure S2A). The lack of effect of codon bias on transgene overexpression is consistent with measurements using synthetic gene libraries in *E. coli*, which detected little correlation between yield of overexpressed proteins and the codon adaptation index (Kudla et al., 2009; Welch et al., 2009). Thus, based on the elongation kinetics inferred from ribosome profiling, we suggest that the major codon bias is not a significant determinant of protein yield during transgene overexpression in *E. coli*.

Differential Aminoacylation Determines the Specificity of Elongation Pausing

Having formulated our biophysical model during nutrient-rich growth when initiation is rate limiting for translation, we sought

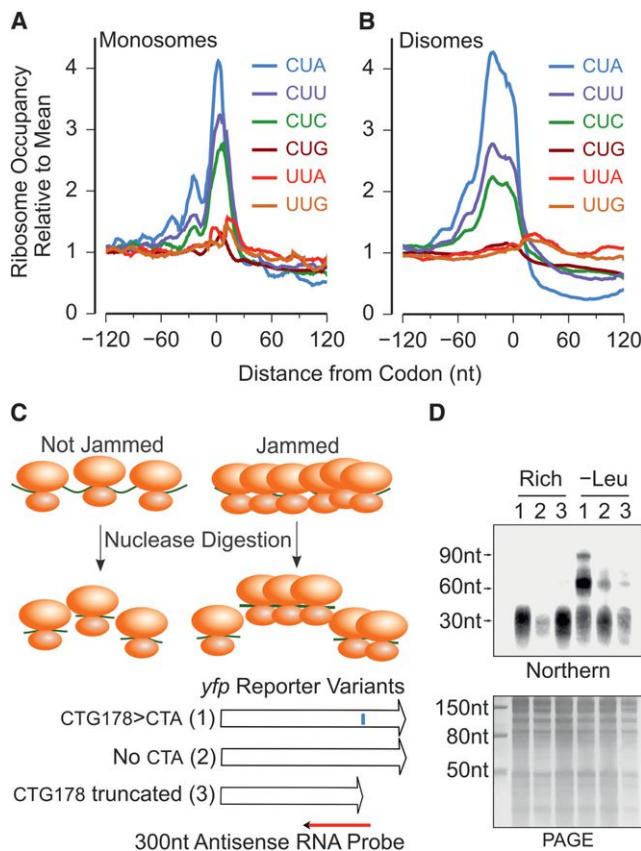


Figure 4. Ribosome Traffic Jams at Ribosome Pause Sites

(A) Measured monosome occupancy from -120 nt to +120 nt around the six leucine codons during leucine starvation. The monosome occupancy was averaged across all occurrences of each codon in the transcriptome.

(B) Measured disome occupancy from -120 nt to +120 nt around the six leucine codons during leucine starvation. The disome occupancy was averaged across all occurrences of each codon in the transcriptome.

(C) Nuclease footprinting assay for detecting ribosome traffic jams on *yfp* reporter mRNAs. The blue vertical bar along the first variant indicates the location of the CTA178>CTA substitution. Northern blotting was performed using a ^{32}P -labeled antisense RNA complementary to the 300 nt mRNA region from -250 nt to +50 nt of the CTA178>CTA substitution.

(D) (Top) Northern blot of nuclease-digested polysomes for the three *yfp* variants. (Bottom) Polyacrylamide gel corresponding to the northern blot. Numbers above individual lanes correspond to the three *yfp* variants in (C). The size markers on the left of the northern blot were inferred by aligning it to the polyacrylamide gel image. The arrows at 30, 60, and 90 nt indicate the approximate locations of monosomes, disomes, and trisomes, respectively.

See also Figure S3.

to test our model in an elongation-limited regime during starvation for single amino acids. To simulate starvation for a single amino acid, we reduced the aminoacylation rate constant of the corresponding tRNA isoacceptors (k^{aa} in Equation 3 in Figure 2B) while keeping all other parameters identical to those during nutrient-rich growth. Reducing the leucylation or the serylization rate constant caused a nonuniform increase in ribosome occupancy at leucine and serine codons, respectively (Figures 3C and S2B). However, the hierarchy predicted by the model significantly devi-

ated from the measured hierarchy among these codons during both leucine and serine starvation (Figures 1B and 1C).

The discrepancy between model and experiments in the hierarchy of ribosome occupancy at codons during starvation could arise from two sources. One possibility is that the uniform rate constants for codon-tRNA interaction (k^{el}) could differ between tRNA isoacceptors, as we used only average estimates in our model. However, when we replaced the average estimates with in vitro measurements of k^{el} , which vary over a 3-fold range for the leucine family (Sørensen et al., 2005), it had little effect on the predicted hierarchy of ribosome occupancy at codons (Figure S2C). A second explanation for the above discrepancy is that the rate constants for aminoacylation (k^{aa}) might be unequal between the different tRNA isoacceptors. Indeed, when we allowed the aminoacylation rate constants for tRNA isoacceptors to differ over a 3-fold range (Figure S2D), the model could largely recapitulate the experimentally measured hierarchy among codons during starvation (Figures 3D and S2E). Our modeling of differential aminoacylation kinetics is consistent with previous in vitro measurements, which found that tRNA isoacceptors could differ 2- to 20-fold in their aminoacylation rate constants (Fender et al., 2004; Harris and Marashi, 1980; Myers et al., 1971).

Accounting for aminoacylation itself has a critical effect on protein expression during amino acid starvation. A recent modeling study arrived at the conclusion that protein expression can be rescued during amino acid starvation by reducing the number of ribosomes in the cell (Shah et al., 2013). This conclusion was based on lowering the total number of tRNAs to simulate amino acid starvation without considering aminoacylation (Figure S2F). However, there is little evidence that the total concentration of tRNAs significantly changes upon amino acid starvation, whereas the canonical effect of amino acid starvation is the reduction in aminoacyl-tRNA concentration due to lower aminoacylation rate (Dittmar et al., 2005; Sørensen, 2001; Sørensen et al., 2005). When we simulated starvation by reducing the levels of aminoacyl-tRNA rather than that of total tRNA using the model of Shah et al., (2013), we did not observe rescue of protein expression upon decreasing the number of ribosomes (Figure S2G). Thus, accounting for aminoacylation qualitatively alters the prediction from biophysical models of translation in the elongation-limited regime of amino acid starvation.

Ribosome Traffic Jams at Ribosome Pause Sites during Amino Acid Starvation

Single mRNAs are often simultaneously translated by several ribosomes. As a result, if ribosomes pause for a sufficiently long duration during elongation, a traffic jam of trailing ribosomes can occur in the 5' region of the ribosome pause site. To detect traffic jams in our ribosome profiling measurement, we calculated the average ribosome occupancy across the transcriptome in the 120 nt region on either side of the leucine codons during leucine starvation (Figure 4A). We observed an increase in ribosome occupancy at three of the leucine codons—CUA, CUC, and CUU—which is consistent with ribosomes pausing at these codons during elongation. In addition, we observed smaller peaks in ribosome density centered approximately -28 nt and -56 nt upstream of these three codons (Figure 4A). These peaks are consistent with a traffic jam of one and two ribosomes behind

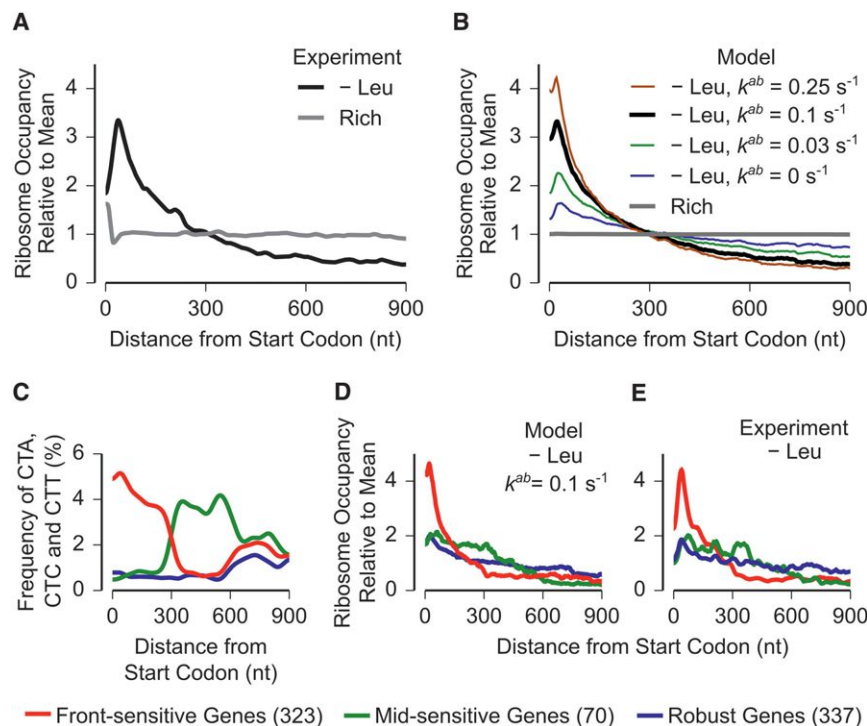


Figure 5. Translation Abortion Determines the Distribution of Ribosomes along mRNAs during Amino Acid Starvation

(A) Measured ribosome occupancy along mRNAs averaged across the transcriptome (1,518 genes). (B) Ribosome occupancy along mRNAs averaged across the transcriptome (1,518 genes) calculated from the whole-cell model. The abortion rate constant k^{ab} was varied. Leucine starvation was modeled as a constant 100-fold reduction in the leucylation rate constant, $k^{aa, \text{Leu}}$.

(C) Codon frequency of the three leucine codons—CTA, CTC, and CTT—in three sets of genes (red, green, blue) with different intragenic distributions of these codons. The number of genes in each class is shown between parentheses in the legend. The codon frequency distribution was smoothed using a Gaussian window of 30 nt width.

(D) Ribosome occupancy averaged across the three sets of genes during leucine starvation calculated from the whole-cell model.

(E) Measured ribosome occupancy averaged across the three sets of genes during leucine starvation.

Ribosome occupancy profiles in all panels were smoothed using a sliding window of 30 nt. Each ribosome occupancy profile was normalized to have a mean value of 1. See also Figure S4.

the paused ribosome. We observed similar but smaller peaks at -28 nt for four of the six serine codons during serine starvation (Figure S3A). Notably, we did not observe smaller peaks upstream of Shine-Dalgarno-like codons, which also cause an increase in ribosome occupancy during nutrient-rich growth (Li et al., 2012) (Figure S3B).

Previous work has shown that nuclease treatment of polysomes with ribosome traffic jams can result in longer mRNA footprints that are protected by multiple ribosomes (Guydosh and Green, 2014; Wolin and Walter, 1988). To test whether longer mRNA footprints occur during amino acid starvation in *E. coli*, we measured the size of nuclease-protected mRNA fragments using three yellow fluorescent protein (*yfp*) reporter variants (Figure 4C). We observed a ~ 30 nt fragment corresponding to the monosome for all *yfp* variants both during rich growth and during leucine starvation (Figure 4D). In addition, we observed a prominent ~ 60 nt fragment and a weaker ~ 90 nt fragment in the *yfp* variant with the CUA codon during leucine starvation. These longer fragments were either faint or absent in the control variants without CUA codons and during nutrient-rich growth, which is consistent with a traffic jam of multiple ribosomes caused by pausing of ribosomes at the CUA codon during leucine starvation. These results also suggest that the standard ribosome profiling method, in which only monosome-protected fragments are sequenced, underestimates the actual in vivo magnitude of ribosome occupancy 5' to the pause site in the presence of ribosome traffic jams. Deep sequencing of longer mRNA fragments (50–80 nt) that were protected by two ribosomes (disomes) from nuclease digestion revealed an increase in average disome occupancy up to 90 nt

upstream of CUA codons during leucine starvation (Figure 4B), which is consistent with a traffic jam of between two to three ribosomes behind the paused ribosome. The length of this region with increased footprint density was similar to that obtained using the standard ribosome profiling method in which only short (~ 30 nt) monosome-protected mRNA fragments were sequenced (Figure 4A), which suggests that both the paused, leading ribosome as well as the jammed, trailing ribosomes have an equal likelihood of occurring either as monosomes or as disomes upon nuclease treatment.

Translation Abortion Determines the Distribution of Ribosomes along mRNAs during Amino Acid Starvation

The strong ribosome pausing that we observed during starvation for single amino acids enabled us to test whether the measured ribosome occupancy along mRNAs is quantitatively consistent with a biophysical model of ribosome traffic jams. During nutrient-rich growth, except for a peak at the start codon, ribosome occupancy was uniformly distributed across the entire length of the mRNA (Figure 5A, gray line), which is similar to previous observations (Oh et al., 2011). By contrast, during starvation for leucine, the distribution of ribosome occupancy was highly nonuniform, varying over a 10-fold range within the first 900 nt (Figure 5A, black line). Ribosome occupancy increased sharply to a maximum at around 30 nt from the start codon and then decreased toward the 3' end of the mRNA. Starvation for serine produced a similar but less skewed distribution of ribosome occupancy along mRNAs (Figure S4A). By comparison, the distribution of total mRNA density was not significantly skewed during leucine starvation (Figure S4B).

We then used the distribution of measured ribosome occupancy along mRNAs to evaluate our biophysical model of translation. Our model reproduced the uniform distribution of measured ribosome occupancy during nutrient-rich growth (Figure 5B, gray line), which is consistent with initiation being rate limiting for translation under these conditions (Li et al., 2014). However, when we simulated leucine starvation by a 100-fold reduction in the leucylation rate constant, a model that included only ribosome traffic jams predicted a more uniform distribution of ribosome occupancy than the one observed experimentally (Figure 5B, blue line, compared to black line in Figure 5A). This disagreement between the model and our measurements could not be offset by varying the leucylation rate, the only remaining free parameter in our model, over a 1,000-fold range (Figure S4C) and was also observed for serine starvation (Figure S4D).

The discrepancy in the distribution of ribosome occupancy between ribosome profiling measurements and our model that included only ribosome traffic jams led us to consider other molecular processes that occur in response to ribosome pausing. Translation abortion is a widespread mechanism of ribosome rescue that has been observed in both bacteria and eukaryotes in response to ribosome pausing (Shoemaker and Green, 2012). Inactivation of translation abortion factors can decrease the growth rate of cells during stressful perturbations (Keiler and Feaga, 2014), indicating the importance of abortion during stress. We modeled translation abortion (Equation 4 in Figure 2B) as occurring at a constant rate k^{ab} whenever the elongation rate of ribosomes falls below a threshold, $R^{el}_{threshold}$. This threshold elongation rate was set below the elongation rate of all codons under nutrient-rich growth to account for the experimental observation that abortion is generally selective for paused ribosomes (Moore and Sauer, 2005). In our model, increasing the rate of translation abortion (k^{ab}) produced higher ribosome occupancy at the 5' end of mRNAs during starvation (Figures 5B and S4D). An in vivo abortion rate (k^{ab}) of $\sim 0.1 \text{ s}^{-1}$ in our model predicted an average distribution of ribosome occupancy that matched the experimentally observed distribution during both leucine and serine starvation (Figures 5B and S4D, black lines).

To test further the role of translation abortion at ribosome pause sites in determining the distribution of ribosome occupancy along mRNAs, we analyzed three subsets of *E. coli* genes with varying codon usage patterns (Figure 5C). These gene sets had different intragenic locations of the three leucine codons CUA, CUC, and CUU that result in ribosome pausing during leucine starvation (Figure 1B). These three codons were present at a high frequency within the first 300 nt in the first set (Figure 5C, red, 323 genes), between 300 and 600 nt in the second set (Figure 5C, green, 70 genes), and at a low frequency throughout the third set (Figure 5C, blue, 337 genes). The whole-cell model incorporating abortion predicted a decreased ribosome occupancy following the mRNA region where the CUA, CUC, and CUU codons were present at high frequency in these gene sets (Figure 5D). This prediction matched the measured distribution of ribosome occupancy for all three of the gene sets (Figure 5E). By contrast, a model that incorporated only ribosome traffic jams did not recapitulate the measured distribution of ribosome occupancy in the three gene sets (Figure S4E). Together, these observations indicate that incorporating abortion at ribo-

some pause sites in our whole-cell model of translation is necessary for quantitative consistency with the measured ribosome occupancy along mRNAs during starvation for single amino acids in *E. coli*.

Translation Abortion and Its Effectors during Amino Acid Starvation

Based on the prediction of translation abortion from our analysis of the measured ribosome occupancy, we looked for abortion events at ribosome pause sites using a *yfp* reporter system. We constructed synonymous variants of *yfp* that encoded a 3xFLAG epitope at the N terminus for detection of incomplete polypeptides and that had a single leucine starvation-sensitive CTA codon at one of three different locations along *3xflag-yfp* (Figure 6A). During leucine starvation, we detected shorter-length polypeptide fragments whose size was consistent with premature abortion at the CUA codon (Figure 6B).

Paused ribosomes with an empty A site are stable in vitro (Ivanova et al., 2005), which suggests a role for a *trans*-acting factor in mediating abortion at ribosome pause sites during leucine starvation. Hence, we tested the role of four known abortion-mediating factors in *E. coli*—tmRNA (Keiler et al., 1996), RF3 (Zaher and Green, 2011), ArfA (Chadani et al., 2010), and ArfB (Chadani et al., 2011)—by deleting the genes for each of these factors individually (Figure 6C). Deletion of the gene encoding *tmRNA*, a chimeric transfer-messenger RNA that releases paused ribosomes and tags the incomplete polypeptide for proteolysis (Keiler et al., 1996), caused a significant increase in the ratio of the incomplete-to-full-length YFP (Figure 6C). Further, complementation with a *tmRNA*_{His6} mutant (Roche and Sauer, 2001), which adds a modified His6 proteolysis-resistant tag but still causes ribosome release, revealed an incomplete YFP polypeptide with a His6 epitope upon immunoprecipitation (Figure 6D). Together, these experiments suggest that tmRNA mediates abortion at ribosome pause sites during leucine starvation. This conclusion is consistent with a previously ascribed role for tmRNA during amino acid starvation (Garza-Sánchez et al., 2008; Li et al., 2008). However, in contrast to these studies and previous in vitro experiments (Ivanova et al., 2004), we did not find evidence for significant mRNA cleavage near the pause site during amino acid starvation (Figure S5A). Furthermore, the measured total mRNA density around the pause site did not indicate appreciable degradation of the region 3' to the pause site (Figure S5B). Finally, ribosome profiling in a $\Delta tmRNA$ strain showed only a slight decrease in ribosome occupancy at the 5' end of mRNAs compared to a *tmRNA*⁺ strain during leucine starvation (Figure S5C). This observation suggests that abortion is still efficient in the absence of tmRNA and is likely mediated by the abortion-mediating factor ArfA. Such an auxiliary role for ArfA in the absence of tmRNA is supported by earlier studies, which found that both the *arfA* mRNA and a truncated but functional form of the ArfA protein are highly stabilized by the deletion of *tmRNA* (Garza-Sánchez et al., 2011) and that double knockout of *arfA* and *tmRNA* is synthetic lethal (Chadani et al., 2010).

We then tested whether ribosome pausing and abortion also affects the expression level of an endogenous *E. coli* protein, the sigma factor RpoS, that is highly upregulated at the

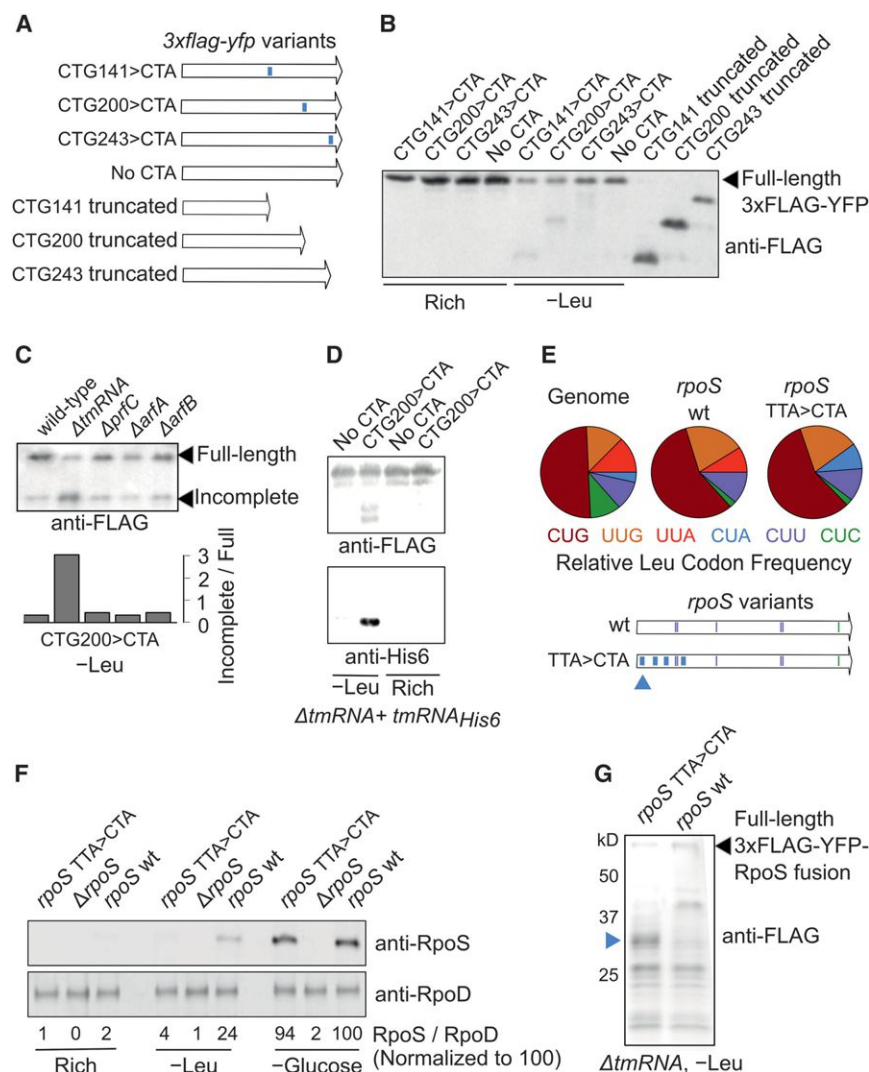


Figure 6. Translation Abortion and Its Effectors during Amino Acid Starvation

(A) Schematic of 3xflag-yfp reporter variants with either single CTA > CTA substitutions (indicated in blue) or truncated at one of three locations.

(B) Western blot using anti-FLAG antibody for the 3xflag-yfp variants shown in (A).

(C) Western blot with anti-FLAG antibody of the CTG200 > CTA variant of yfp during leucine starvation in strains with deletion of one of four different genes encoding factors that mediate translation abortion (*tmRNA*, *prfC*, *arfA*, *arfB*). "Wild-type" refers to the parent leucine auxotroph strain. The bottom panel indicates the densitometric ratio of these two bands.

(D) (Top) Immunoprecipitation with anti-FLAG antibody of CTG200 > CTA yfp variant expressed during leucine starvation in a $\Delta tmRNA$ strain with a *tmRNA_{His6}* mutant. (Bottom) *tmRNA_{His6}* activity detected with an anti-His6 antibody.

(E) (Pie charts) Relative frequency of the six leucine codons across all coding sequences in the genome, in the *rpoS* wild-type coding sequence, and in the *rpoS* synonymous variant. Four TTA codons were replaced by CTA codons in the *rpoS* synonymous variant at the locations indicated by thick blue bars. The thin blue and green bars correspond to the location of the CTC and CTT codons in the *rpoS* WT and synonymous variant. Blue triangle indicates the location of the first ribosome pause site encoded by the CTA codon during leucine starvation.

(F) Western blot against the RpoS protein (top) and RpoD protein (bottom) during nutrient-rich growth, leucine starvation, and glucose starvation. The *rpoS* wild-type coding sequence at the native chromosomal locus was either deleted ($\Delta rpoS$) or replaced by the *rpoS* TTA > CTA synonymous variant without additional selection markers. Numbers between the two panels indicate the normalized densitometric ratio of the RpoS and RpoD bands for each lane.

(G) Western blot with anti-FLAG antibody against 3xFLAG-YFP-RpoS fusion proteins during leucine

starvation. Approximate molecular weight in kilodaltons (kD) was estimated using a protein ladder. Blue triangle corresponds to the expected truncated peptide caused by ribosome abortion at the first pause site in the *rpoS* TTA > CTA synonymous variant (indicated as a blue triangle in E). See also Figure S5 and Table S2.

transcriptional level in response to leucine starvation (Subramaniam et al., 2013a). Ribosome pause sites encoded by the two leucine codons CTA and CTC during leucine starvation are underrepresented in the *rpoS* protein coding sequence in comparison to their average frequency in the genome (Figure 6E), which suggests that ribosome pause sites during leucine starvation are selected against in the *rpoS* coding sequence due to their adverse effect on the expression of this critical stress response protein. To test the effect of ribosome pause sites on RpoS expression, we substituted, at the chromosomal *rpoS* locus, four leucine TTA codons with the synonymous CTA codons that encode ribosome pause sites (Figure 6E, thick blue bars). Western blotting indicated that the expression level of RpoS decreased ~6-fold during leucine starvation due to the presence of the CTA codons in the *rpoS* coding sequence (Figure 6F). This decrease was specific to leucine starvation, as the synonymous

substitutions did not significantly affect the expression of RpoS during glucose starvation, which also strongly upregulates RpoS (Figure 6F). By fusing 3xFLAG-YFP to the N terminus of RpoS, we detected a truncated peptide (Figure 6G, blue triangle) whose size is consistent with translation abortion at the first ribosome pause site in the *rpoS* synonymous variant during leucine starvation (Figure 6E, blue triangle).

Effect of Translation Abortion on Protein Expression during Amino Acid Starvation

Previous studies have suggested that the primary function of abortion mediated by factors such as *tmRNA* is to rescue inactive mRNA-bound ribosomes during stress and thereby increase the translational capacity in the cell (Moore and Sauer, 2007). To quantify this effect of abortion on ribosome rescue and translational capacity, we tracked the number of free ribosomes and

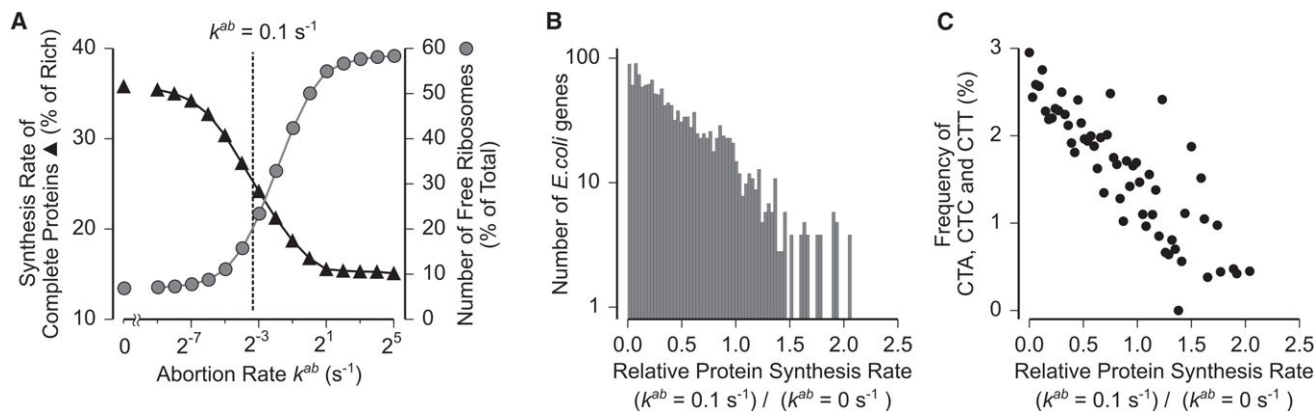


Figure 7. Effect of Translation Abortion on Protein Expression

(A) Effect of varying abortion rate constant (k^{ab}) on the number of free ribosomes in the cell (gray circles) and the global synthesis rate of complete proteins (black triangles) during leucine starvation calculated from whole-cell model. The value $k^{ab} = 0.1 \text{ s}^{-1}$ that fits the measured ribosome occupancy (black line in Figure 5B) is indicated as a dashed line.

(B) Effect of non-zero abortion rate constant on the synthesis rate of individual *E. coli* proteins during leucine starvation calculated from the whole-cell model.

(C) Average frequency of the three leucine codons CTA, CTC, and CTT for genes in each of the histogram bins in (B). Only genes with greater than ten leucine codons were considered in (B) and (C).

See also Figure S6.

the global synthesis rate of proteins as a function of the abortion rate in our whole-cell model (Figures 7A and S6). Based on earlier work (Bremer and Dennis, 2008), we assumed that 15% of the ribosomes in a cell are free (not bound to mRNAs) during nutrient-rich growth. In the absence of abortion ($k^{ab} = 0 \text{ s}^{-1}$), starvation for leucine decreased the fraction of free ribosomes in the cell to 7% (Figure 7A, leftmost circle). Increasing the rate of abortion during leucine starvation gradually rescued the fraction of free ribosomes to a maximum of $\sim 60\%$ in our model (Figure 7A, rightmost circle). We then examined the effect of abortion on global protein expression using our whole-cell model. In the absence of abortion ($k^{ab} = 0 \text{ s}^{-1}$), leucine starvation decreased the synthesis rate of complete proteins in the cell to 35% of its value during nutrient-rich growth (Figure 7A, leftmost triangle). Increasing the rate of abortion during leucine starvation further decreased the global synthesis rate to a minimum of $\sim 15\%$ (Figure 7A, rightmost triangle). Notably, increasing the abortion rate had widely different effects on protein expression from individual mRNAs during leucine starvation (Figure 7B). Proteins with higher expression level upon increasing the abortion rate also had a lower frequency of ribosome pause sites in the corresponding mRNAs (Figure 7C). As a result, protein expression from these mRNAs is less susceptible to abortion, and their translation initiation rate is increased by the released ribosomes due to abortion from other mRNAs. Thus, our model predicts that abortion enhances the translation of mRNAs in a selective manner even though it can have a deleterious effect on global protein expression during amino acid starvation.

DISCUSSION

Here, we formulated a biophysical model of translation in *E. coli* by leveraging the near single-codon resolution of the ribosome profiling method. Notably, we constrained our model not only

in the initiation-limited regime of nutrient-rich growth, but also in the regime of amino acid starvation, during which the elongation rate of ribosomes has a large effect on the measured ribosome occupancy. By contrast, previous computational models of translation used data solely from initiation-limited regimes, with the consequence that mechanistic features of the elongation stage of translation were not fully constrained (Shah et al., 2013; Tuller et al., 2010). As our results generally illustrate, the exact mechanistic features assumed for the elongation stage have a critical role in models of translation, and changing these features qualitatively alters several of the model predictions.

Consistency between ribosome profiling measurements and our model suggests that the concentration of aminoacyl-tRNAs does not limit the elongation rate of ribosomes at most codons during nutrient-rich growth. This conclusion relies on our assumption that biases in the generation and analysis of ribosome footprinting data are sufficiently small such that the measured ribosome occupancy at codons reflects the residence times of ribosomes with an empty A site at these codons. This assumption is partly supported by our observation that differences in ribosome occupancy at codons during starvation for their cognate amino acid are consistent with their corresponding effect on protein expression. A more direct test of our conclusion will be to characterize the effect of overexpressing low-abundance tRNA isoacceptors on the measured ribosome occupancy during nutrient-rich growth. Our model still predicts a small variation in ribosome occupancy at codons that is correlated with tRNA abundance, the exact magnitude of which depends on the k_{cat}/K_M for ribosome-tRNA association (Figure 3A, vertical axis). Although we predict that this small variation neither limits the ribosome elongation rate at most codons nor affects protein levels during nutrient-rich growth on the physiological timescale of protein synthesis, it can nevertheless leave an evolutionary signature on protein coding sequences and thus underlie the

widely observed correlation between tRNA abundance and codon frequency in microorganisms (Andersson and Kurland, 1990; Drummond and Wilke, 2008; Wallace et al., 2013).

Our analysis of aminoacylation kinetics raises the interesting possibility that cells might utilize differential aminoacylation rates of tRNA isoacceptors as a mechanism to regulate the elongation rate of codons specifically during nutrient stress. Further, small differences in aminoacylation rate between tRNA isoacceptors might also modulate mis-aminacylation rates during other stresses (Netzer et al., 2009) while having little or no deleterious effect on translation during nutrient-rich growth when aminoacylation is not limiting for elongation.

Our work provides *in vivo* evidence for the widespread occurrence of ribosome traffic jams in response to pausing of a leading ribosome. However, we found that translation abortion at ribosome pause sites is the primary determinant of ribosome occupancy along mRNAs during amino acid starvation. Although abortion from non-stop mRNAs is thought to increase global protein expression during nutrient-rich growth (Moore and Sauer, 2007), our whole-cell modeling suggests that abortion of paused ribosomes during nutrient stress might have the function of enabling selective translation from specific mRNAs. Additionally, releasing nascent polypeptides from paused ribosomes by abortion might facilitate their proteolysis and prevent protein misfolding during stress. However, previous studies have found that eliminating the ability of the abortion-mediating factor tmRNA to target nascent polypeptides to proteolysis does not significantly alter the cellular response to stress (Abo et al., 2002; Huang et al., 2000), suggesting that prevention of protein misfolding is not the primary cellular function of abortion-mediating factors during stress.

The four processes that we modeled in our study (Figure 2) are an essential part of protein synthesis in both bacteria and eukaryotes. Hence, our analysis in *E. coli* can be readily extended to eukaryotes, where ribosome profiling has revealed that a variety of stresses result in a slow elongation rate (Liu et al., 2013; Shalgi et al., 2013). It will also be useful to integrate our whole-cell model of translation with quantitative models of other cellular processes such as transcription (Brewster et al., 2014) and metabolism (Bordbar et al., 2014). Such an integrated model can shed light on the complex interplay between metabolism and gene expression that occurs during environmental changes (Subramaniam et al., 2013b).

EXPERIMENTAL PROCEDURES

Construction of all strains and plasmids (listed in Table S2), western blots, and northern blots were performed using standard molecular biology techniques (Extended Experimental Procedures).

Ribosome Profiling

Ribosome profiling was carried out as described previously (Li et al., 2012; Oh et al., 2011), with the following modifications. To accurately capture the ribosome occupancy on mRNAs with single-codon resolution, we flash froze the cells immediately upon harvesting and stabilized ribosomes with the translation inhibitor chloramphenicol only at the lysis stage. Cells were lysed using glass beads (G1277, Sigma, vortex 10 × 30 s at 4°C with 60 s cooling on ice in between). Micrococcal nuclease digestion was carried out with 1 U Worthington Biochemicals MNase per μg of nucleic acid, as measured by A_{260} . Monosome-protected mRNA footprints between 20 and 40 nt were size

selected by polyacrylamide gel electrophoresis for monosome sequencing. For disome sequencing, the disome peak was collected from the MNase-treated polysomes after sucrose-gradient fractionation, and fragments between 50 and 80 nt were used for sequencing. For total mRNA sequencing, the Microbe Express kit (Ambion) was used for subtracting rRNA from total mRNA and then fragmented using a bicarbonate buffer (Ingolia et al., 2009) for 20 min. For library construction, polyA-tailing (Ingolia et al., 2009) was used instead of linker ligation.

High-Throughput Sequencing Data Analysis

Analysis steps were similar to that in previous ribosome profiling studies (Li et al., 2012; Subramaniam et al., 2013b) and were implemented using Python and Bash programming languages. Full programming code for generating the final figures in our paper starting from raw sequencing data is provided both as an interactive IPython notebook (Perez and Granger, 2007) and as a static HTML file (sequencing_data_analysis.html in Data S1). In brief, single-end reads were polyA trimmed and then aligned to the *E. coli* genome (NC_000913.3 build) using Bowtie (Langmead et al., 2009). Aligned reads were trimmed by 8 nt on each side. Each genomic position corresponding to the trimmed read was assigned ribosome occupancy equal to the inverse of the read length. The transcriptome-averaged ribosome occupancy at individual codons for each sample (Figures 1B and 1C) was computed by first averaging the ribosome occupancy at the first nucleotide position of the codon across all occurrences of that codon within each coding sequence and then by averaging across all coding sequences that had a minimum average occupancy of one read/codon. Because the start codon and the three stop codons have a high ribosome density during both nutrient-rich growth and starvation, they were excluded in the plots showing all 61 sense codons (Figures 1 and S1) for clarity. The ribosome occupancy profiles around codons (Figures 5A and 5B) were calculated using the same procedure but for the 120 nt region on each side of the codon rather than just the first nucleotide position of the codon. Ribosome occupancy along mRNAs (Figure 6A) was computed by first normalizing the ribosome occupancy at each position of a coding sequence by the average ribosome occupancy for the full coding sequence, and then by averaging this quantity at each position beginning from the start codon across all coding sequences that had a minimum average occupancy of one read/codon.

Whole-Cell Model of *E. coli* Translation

Our whole-cell model for translation (Figure 2) was implemented using the stochastic Gillespie algorithm (Gillespie, 1977). The source code for implementing the kinetic model in Figure 2 was adapted from Shah et al., (2013), with modifications as shown in Data S1. Full programming code for running the simulation and for the subsequent data analysis to generate the final figures in our paper is provided both as an interactive IPython notebook (Perez and Granger, 2007) and as a static HTML file (simulation.html and simulation_data_analysis.html in Data S1). Standard errors of mean for model predictions in all figures are smaller than data markers. Our whole-cell model tracked the state of 44,000 ribosomes, 408,000 tRNA molecules (38 distinct tRNA species), and ~7,500 mRNA molecules (1,518 distinct mRNA species) in the *E. coli* cell (Table S1). The transition rates between different states of each of these molecules are determined by the four rate equations in Figure 2B. The key parameters that control the predictions from our model are the rate constants k for the four processes in Figure 2B. Based on the analysis presented in the main text, we chose a default set of parameters for all of our simulations. Table S1 lists these choices, together with the corresponding references and a footnote explaining the choice (see also Extended Experimental Procedures). We chose parameters corresponding to a cell doubling time of 30 min that we measured during nutrient-rich growth in our experiments.

ACCESSION NUMBERS

Sequencing data are publicly available from Gene Expression Omnibus, accession number GSE51052. Simulation results and programming code for simulation, sequencing data analysis, simulation data analysis and reproduction of figures are publicly available at <http://datadryad.org/>, <http://dx.doi.org/10.5061/dryad.ch352>.

SUPPLEMENTAL INFORMATION

Supplemental Information includes Extended Experimental Procedures, six figures, two tables, and one data file and can be found with this article online at <http://dx.doi.org/10.1016/j.cell.2014.10.043>.

AUTHOR CONTRIBUTIONS

A.R.S. and E.K.O. designed research. A.R.S. performed research and analyzed data with input from B.M.Z. A.R.S. and E.K.O. wrote the manuscript.

ACKNOWLEDGMENTS

We thank J. Calarco, P. Cluzel, V. Denic, A. Murray, and C. Shoemaker for discussions; G.W. Li and E. Oh for advice on ribosome profiling; G.L. Chew and A. Gutu for advice on northern blotting; H. Aiba for the XL001 strain; P. Shah for clarifications on the yeast model; and K. Amarnath, A. Buskirk, S. Chandrasekaran, A. Darnell, G.W. Li, W. Moebius, J. Moffitt, P. Shah, and E. Wallace for comments on the manuscript. The computations in this paper were run on the Odyssey cluster supported by the FAS Division of Science, Research Computing Group at Harvard University. This research was supported by an NIH K99 Pathway to Independence Award GM107113 (A.R.S.) and by the Howard Hughes Medical Institute (E.K.O.).

Received: June 16, 2014

Revised: August 18, 2014

Accepted: October 21, 2014

Published: November 20, 2014

REFERENCES

- Abo, T., Ueda, K., Sunohara, T., Ogawa, K., and Aiba, H. (2002). SsrA-mediated protein tagging in the presence of miscoding drugs and its physiological role in *Escherichia coli*. *Genes Cells* 7, 629–638.
- Andersson, S.G., and Kurland, C.G. (1990). Codon preferences in free-living microorganisms. *Microbiol. Rev.* 54, 198–210.
- Bilgin, N., and Ehrenberg, M. (1994). Mutations in 23 S ribosomal RNA perturb transfer RNA selection and can lead to streptomycin dependence. *J. Mol. Biol.* 235, 813–824.
- Bordbar, A., Monk, J.M., King, Z.A., and Palsson, B.O. (2014). Constraint-based models predict metabolic and associated cellular functions. *Nat. Rev. Genet.* 15, 107–120.
- Bremer, H., and Dennis, P. (2008). Modulation of chemical composition and other parameters of the cell at different exponential growth rates. In *EcoSal-*Escherichia coli* and *Salmonella*: Cellular and Molecular Biology*, A. Bock, R. Curtiss, III, J.B. Kaper, P.D. Karp, F.C. Neidhardt, T. Nystrom, J.M. Slauch, C.L. Squires, D. Ussery, and E. Schaechter, eds. (Washington, DC: ASM Press).
- Brewster, R.C., Weinert, F.M., Garcia, H.G., Song, D., Rydenfelt, M., and Phillips, R. (2014). The transcription factor titration effect dictates level of gene expression. *Cell* 156, 1312–1323.
- Chadani, Y., Ono, K., Ozawa, S., Takahashi, Y., Takai, K., Nanamiya, H., Tozawa, Y., Kutsukake, K., and Abo, T. (2010). Ribosome rescue by *Escherichia coli* ArfA (YhdL) in the absence of trans-translation system. *Mol. Microbiol.* 78, 796–808.
- Chadani, Y., Ono, K., Kutsukake, K., and Abo, T. (2011). *Escherichia coli* YaeJ protein mediates a novel ribosome-rescue pathway distinct from SsrA- and ArfA-mediated pathways. *Mol. Microbiol.* 80, 772–785.
- Dittmar, K.A., Sorensen, M.A., Elf, J., Ehrenberg, M., and Pan, T. (2005). Selective charging of tRNA isoacceptors induced by amino-acid starvation. *EMBO Rep.* 6, 151–157.
- Dong, H., Nilsson, L., and Kurland, C.G. (1996). Co-variation of tRNA abundance and codon usage in *Escherichia coli* at different growth rates. *J. Mol. Biol.* 260, 649–663.
- Drummond, D.A., and Wilke, C.O. (2008). Mistranslation-induced protein misfolding as a dominant constraint on coding-sequence evolution. *Cell* 134, 341–352.
- Elf, J., Nilsson, D., Tenson, T., and Ehrenberg, M. (2003). Selective charging of tRNA isoacceptors explains patterns of codon usage. *Science* 300, 1718–1722.
- Fender, A., Sissler, M., Florentz, C., and Giegé, R. (2004). Functional idiosyncrasies of tRNA isoacceptors in cognate and noncognate aminoacylation systems. *Biochimie* 86, 21–29.
- Garza-Sánchez, F., Gin, J.G., and Hayes, C.S. (2008). Amino acid starvation and colicin D treatment induce A-site mRNA cleavage in *Escherichia coli*. *J. Mol. Biol.* 378, 505–519.
- Garza-Sánchez, F., Schaub, R.E., Janssen, B.D., and Hayes, C.S. (2011). tmRNA regulates synthesis of the ArfA ribosome rescue factor. *Mol. Microbiol.* 80, 1204–1219.
- Gillespie, D.T. (1977). Exact stochastic simulation of coupled chemical reactions. *J. Phys. Chem.* 81, 2340–2361.
- Gustafsson, C., Govindarajan, S., and Minshull, J. (2004). Codon bias and heterologous protein expression. *Trends Biotechnol.* 22, 346–353.
- Guydosh, N.R., and Green, R. (2014). Dom34 rescues ribosomes in 3' untranslated regions. *Cell* 156, 950–962.
- Harris, C.L., and Marashi, F. (1980). Two kinetically distinct tRNA^{Ala} isoacceptors in *Escherichia coli* C6. *Nucleic Acids Res.* 8, 2023–2037.
- Huang, C., Wolfgang, M.C., Withey, J., Koomey, M., and Friedman, D.I. (2000). Charged tmRNA but not tmRNA-mediated proteolysis is essential for *Neisseria gonorrhoeae* viability. *EMBO J.* 19, 1098–1107.
- Ingolia, N.T., Ghaemmaghami, S., Newman, J.R.S., and Weissman, J.S. (2009). Genome-wide analysis in vivo of translation with nucleotide resolution using ribosome profiling. *Science* 324, 218–223.
- Ivanova, N., Pavlov, M.Y., Felden, B., and Ehrenberg, M. (2004). Ribosome rescue by tmRNA requires truncated mRNAs. *J. Mol. Biol.* 338, 33–41.
- Ivanova, N., Pavlov, M.Y., and Ehrenberg, M. (2005). tmRNA-induced release of messenger RNA from stalled ribosomes. *J. Mol. Biol.* 350, 897–905.
- Jacques, N., and Dreyfus, M. (1990). Translation initiation in *Escherichia coli*: old and new questions. *Mol. Microbiol.* 4, 1063–1067.
- Keiler, K.C., and Feaga, H.A. (2014). Resolving nonstop translation complexes is a matter of life or death. *J. Bacteriol.* 196, 1490–1494.
- Keiler, K.C., Waller, P.R.H., and Sauer, R.T. (1996). Role of a peptide tagging system in degradation of proteins synthesized from damaged messenger RNA. *Science* 271, 990–993.
- Kudla, G., Murray, A.W., Tollervey, D., and Plotkin, J.B. (2009). Coding-sequence determinants of gene expression in *Escherichia coli*. *Science* 324, 255–258.
- Langmead, B., Trapnell, C., Pop, M., and Salzberg, S.L. (2009). Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol.* 10, R25.
- Li, X., Yagi, M., Morita, T., and Aiba, H. (2008). Cleavage of mRNAs and role of tmRNA system under amino acid starvation in *Escherichia coli*. *Mol. Microbiol.* 68, 462–473.
- Li, G.-W., Oh, E., and Weissman, J.S. (2012). The anti-Shine-Dalgarno sequence drives translational pausing and codon choice in bacteria. *Nature* 484, 538–541.
- Li, G.-W., Burkhardt, D., Gross, C., and Weissman, J.S. (2014). Quantifying absolute protein synthesis rates reveals principles underlying allocation of cellular resources. *Cell* 157, 624–635.
- Liu, B., Han, Y., and Qian, S.-B. (2013). Cotranslational response to proteotoxic stress by elongation pausing of ribosomes. *Mol. Cell* 49, 453–463.
- Moore, S.D., and Sauer, R.T. (2005). Ribosome rescue: tmRNA tagging activity and capacity in *Escherichia coli*. *Mol. Microbiol.* 58, 456–466.
- Moore, S.D., and Sauer, R.T. (2007). The tmRNA system for translational surveillance and ribosome rescue. *Annu. Rev. Biochem.* 76, 101–124.

- Myers, G., Blank, H.U., and Söll, D. (1971). A comparative study of the interactions of *Escherichia coli* leucyl-, seryl-, and valyl-transfer ribonucleic acid synthetases with their cognate transfer ribonucleic acids. *J. Biol. Chem.* **246**, 4955–4964.
- Netzer, N., Goodenbour, J.M., David, A., Dittmar, K.A., Jones, R.B., Schneider, J.R., Boone, D., Eves, E.M., Rosner, M.R., Gibbs, J.S., et al. (2009). Innate immune and chemically triggered oxidative stress modifies translational fidelity. *Nature* **462**, 522–526.
- Oh, E., Becker, A.H., Sandikci, A., Huber, D., Chaba, R., Gloge, F., Nichols, R.J., Typas, A., Gross, C.A., Kramer, G., et al. (2011). Selective ribosome profiling reveals the cotranslational chaperone action of trigger factor in vivo. *Cell* **147**, 1295–1308.
- Pavlov, M.Y., and Ehrenberg, M. (1996). Rate of translation of natural mRNAs in an optimized in vitro system. *Arch. Biochem. Biophys.* **328**, 9–16.
- Perez, F., and Granger, B.E. (2007). IPython: A System for Interactive Scientific Computing. *Comput. Sci. Eng.* **9**, 21–29.
- Qian, W., Yang, J.-R., Pearson, N.M., Maclean, C., and Zhang, J. (2012). Balanced codon usage optimizes eukaryotic translational efficiency. *PLoS Genet.* **8**, e1002603.
- Roche, E.D., and Sauer, R.T. (2001). Identification of endogenous SsrA-tagged proteins reveals tagging at positions corresponding to stop codons. *J. Biol. Chem.* **276**, 28509–28515.
- Shah, P., Ding, Y., Niemczyk, M., Kudla, G., and Plotkin, J.B. (2013). Rate-limiting steps in yeast protein translation. *Cell* **153**, 1589–1601.
- Shalgi, R., Hurt, J.A., Krykbaeva, I., Taipale, M., Lindquist, S., and Burge, C.B. (2013). Widespread regulation of translation by elongation pausing in heat shock. *Mol. Cell* **49**, 439–452.
- Shoemaker, C.J., and Green, R. (2012). Translation drives mRNA quality control. *Nat. Struct. Mol. Biol.* **19**, 594–601.
- Sørensen, M.A. (2001). Charging levels of four tRNA species in *Escherichia coli* Rel(+) and Rel(-) strains during amino acid starvation: a simple model for the effect of ppGpp on translational accuracy. *J. Mol. Biol.* **307**, 785–798.
- Sørensen, M.A., Elf, J., Bouakaz, E., Tenson, T., Sanyal, S., Björk, G.R., and Ehrenberg, M. (2005). Over expression of a tRNA(Leu) isoacceptor changes charging pattern of leucine tRNAs and reveals new codon reading. *J. Mol. Biol.* **354**, 16–24.
- Subramaniam, A.R., Pan, T., and Cluzel, P. (2013a). Environmental perturbations lift the degeneracy of the genetic code to regulate protein levels in bacteria. *Proc. Natl. Acad. Sci. USA* **110**, 2419–2424.
- Subramaniam, A.R., DeLoughery, A., Bradshaw, N., Chen, Y., O'Shea, E., Losick, R., and Chai, Y. (2013b). A serine sensor for multicellularity in a bacterium. *eLife* **2**, e01501.
- Tuller, T., Carmi, A., Vestsigian, K., Navon, S., Dorfan, Y., Zaborske, J., Pan, T., Dahan, O., Furman, I., and Pilpel, Y. (2010). An evolutionarily conserved mechanism for controlling the efficiency of protein translation. *Cell* **141**, 344–354.
- Varenne, S., Buc, J., Llobes, R., and Lazdunski, C. (1984). Translation is a non-uniform process. Effect of tRNA availability on the rate of elongation of nascent polypeptide chains. *J. Mol. Biol.* **180**, 549–576.
- Vind, J., Sørensen, M.A., Rasmussen, M.D., and Pedersen, S. (1993). Synthesis of proteins in *Escherichia coli* is limited by the concentration of free ribosomes. Expression from reporter genes does not always reflect functional mRNA levels. *J. Mol. Biol.* **231**, 678–688.
- Wallace, E.W.J., Airolidi, E.M., and Drummond, D.A. (2013). Estimating selection on synonymous codon usage from noisy experimental data. *Mol. Biol. Evol.* **30**, 1438–1453.
- Welch, M., Govindarajan, S., Ness, J.E., Villalobos, A., Gurney, A., Minshull, J., and Gustafsson, C. (2009). Design parameters to control synthetic gene expression in *Escherichia coli*. *PLoS ONE* **4**, e7002.
- Wintermeyer, W., Peske, F., Beringer, M., Gromadski, K.B., Savelsbergh, A., and Rodnina, M.V. (2004). Mechanisms of elongation on the ribosome: dynamics of a macromolecular machine. *Biochem. Soc. Trans.* **32**, 733–737.
- Wolin, S.L., and Walter, P. (1988). Ribosome pausing and stacking during translation of a eukaryotic mRNA. *EMBO J.* **7**, 3559–3569.
- Yegian, C.D., Stent, G.S., and Martin, E.M. (1966). Intracellular condition of *Escherichia coli* transfer RNA. *Proc. Natl. Acad. Sci. USA* **55**, 839–846.
- Zaher, H.S., and Green, R. (2011). A primary role for release factor 3 in quality control during translation elongation in *Escherichia coli*. *Cell* **147**, 396–408.
- Zhang, S., Goldman, E., and Zubay, G. (1994). Clustering of low usage codons and ribosome movement. *J. Theor. Biol.* **170**, 339–354.

A Proteome-Scale Map of the Human Interactome Network

Thomas Rolland,^{1,2,19} Murat Taşan,^{1,3,4,5,19} Benoit Charlotiaux,^{1,2,19} Samuel J. Pevzner,^{1,2,6,7,19} Quan Zhong,^{1,2,8,19} Nidhi Sahni,^{1,2,19} Song Yi,^{1,2,19} Irma Lemmens,⁹ Celia Fontanillo,¹⁰ Roberto Mosca,¹¹ Atanas Kamburov,^{1,2} Susan D. Ghiassian,^{1,12} Xinping Yang,^{1,2} Lila Ghamsari,^{1,2} Dawit Balcha,^{1,2} Bridget E. Begg,^{1,2} Pascal Braun,^{1,2} Marc Brehme,^{1,2} Martin P. Broly,^{1,2} Anne-Ruxandra Carvunis,^{1,2} Dan Convery-Zupan,^{1,2} Roser Corominas,¹³ Jasmin Coulombe-Huntington,^{1,14} Elizabeth Dann,^{1,2} Matija Dreze,^{1,2} Amélie Dricot,^{1,2} Changyu Fan,^{1,2} Eric Franzosa,^{1,14} Fana Gebreab,^{1,2} Bryan J. Gutierrez,^{1,2} Madeleine F. Hardy,^{1,2} Mike Jin,^{1,2} Shuli Kang,¹³ Ruth Kiros,^{1,2} Guan Ning Lin,¹³ Katja Luck,^{1,2} Andrew MacWilliams,^{1,2} Jörg Menche,^{1,12} Ryan R. Murray,^{1,2} Alexandre Palagi,^{1,2} Matthew M. Poulin,^{1,2} Xavier Rambout,^{1,2,15} John Rasla,^{1,2} Patrick Reichert,^{1,2} Viviana Romero,^{1,2} Elien Ruysinck,⁹ Julie M. Sahalie,^{1,2} Annemarie Scholz,^{1,2} Akash A. Shah,^{1,2} Amitabh Sharma,^{1,12} Yun Shen,^{1,2} Kerstin Spirohn,^{1,2} Stanley Tam,^{1,2} Alexander O. Tejeda,^{1,2} Shelly A. Trigg,^{1,2} Jean-Claude Twizere,^{1,2,15} Kerwin Vega,^{1,2} Jennifer Walsh,^{1,2} Michael E. Cusick,^{1,2} Yu Xia,^{1,14} Albert-László Barabási,^{1,12,16} Lilia M. Iakoucheva,¹³ Patrick Aloy,^{11,17} Javier De Las Rivas,¹⁰ Jan Tavernier,⁹ Michael A. Calderwood,^{1,2,20} David E. Hill,^{1,2,20} Tong Hao,^{1,2,20} Frederick P. Roth,^{1,3,4,5,18,*} and Marc Vidal^{1,2,*}

¹Center for Cancer Systems Biology (CCSB) and Department of Cancer Biology, Dana-Farber Cancer Institute, Boston, MA 02215, USA

²Department of Genetics, Harvard Medical School, Boston, MA 02115, USA

³Departments of Molecular Genetics and Computer Science, University of Toronto, Toronto, ON M5S 3E1, Canada

⁴Donnelly Centre, University of Toronto, Toronto, ON M5S 3E1, Canada

⁵Lunenfeld-Tanenbaum Research Institute, Mount Sinai Hospital, Toronto, ON M5G 1X5, Canada

⁶Department of Biomedical Engineering, Boston University, Boston, MA 02215, USA

⁷Boston University School of Medicine, Boston, MA 02118, USA

⁸Department of Biological Sciences, Wright State University, Dayton, OH 45435, USA

⁹Department of Medical Protein Research, VIB, 9000 Ghent, Belgium

¹⁰Cancer Research Center (Centro de Investigación del Cáncer), University of Salamanca and Consejo Superior de Investigaciones Científicas, Salamanca 37008, Spain

¹¹Joint IRB-BSC Program in Computational Biology, Institute for Research in Biomedicine (IRB Barcelona), Barcelona 08028, Spain

¹²Center for Complex Network Research (CCNR) and Department of Physics, Northeastern University, Boston, MA 02115, USA

¹³Department of Psychiatry, University of California, San Diego, La Jolla, CA 92093, USA

¹⁴Department of Bioengineering, McGill University, Montreal, QC H3A 0C3, Canada

¹⁵Protein Signaling and Interactions Lab, GIGA-R, University of Liege, 4000 Liege, Belgium

¹⁶Department of Medicine, Brigham and Women's Hospital, Harvard Medical School, Boston, MA 02115, USA

¹⁷Institució Catalana de Recerca i Estudis Avançats (ICREA), Barcelona 08010, Spain

¹⁸Canadian Institute for Advanced Research, Toronto M5G 1Z8, Canada

¹⁹Co-first author

²⁰Co-senior author

*Correspondence: fritz.roth@utoronto.ca (F.P.R.), marc_vidal@dfci.harvard.edu (M.V.)

<http://dx.doi.org/10.1016/j.cell.2014.10.050>

SUMMARY

Just as reference genome sequences revolutionized human genetics, reference maps of interactome networks will be critical to fully understand genotype-phenotype relationships. Here, we describe a systematic map of ~14,000 high-quality human binary protein-protein interactions. At equal quality, this map is ~30% larger than what is available from small-scale studies published in the literature in the last few decades. While currently available information is highly biased and only covers a relatively small portion of the proteome, our systematic map appears strikingly more homogeneous, revealing a “broader” human interactome network than currently appreciated. The map also uncovers significant interconnectivity between known and candidate cancer

gene products, providing unbiased evidence for an expanded functional cancer landscape, while demonstrating how high-quality interactome models will help “connect the dots” of the genomic revolution.

INTRODUCTION

Since the release of a high-quality human genome sequence a decade ago (International Human Genome Sequencing Consortium, 2004), our ability to assign genotypes to phenotypes has exploded. Genes have been identified for most Mendelian disorders (Hamosh et al., 2005) and over 100,000 alleles have been implicated in at least one disorder (Stenson et al., 2014). Hundreds of susceptibility loci have been uncovered for numerous complex traits (Hindorf et al., 2009) and the genomes of a few thousand human tumors have been nearly fully sequenced (Chin et al., 2011). This genomic revolution is poised to generate a complete description of all relevant genotypic variations in the human population.

Genomic sequencing will, however, if performed in isolation, leave fundamental questions pertaining to genotype-phenotype relationships unresolved (Vidal et al., 2011). The causal changes that connect genotype to phenotype remain generally unknown, especially for complex trait loci and cancer-associated mutations. Even when identified, it is often unclear how a causal mutation perturbs the function of the corresponding gene or gene product. To “connect the dots” of the genomic revolution, functions and context must be assigned to large numbers of genotypic changes.

Complex cellular systems formed by interactions among genes and gene products, or interactome networks, appear to underlie most cellular functions (Vidal et al., 2011). Thus, a full understanding of genotype-phenotype relationships in human will require mechanistic descriptions of how interactome networks are perturbed as a result of inherited and somatic disease susceptibilities. This, in turn, will require high-quality and extensive genome and proteome-scale maps of macromolecular interactions such as protein-protein interactions (PPIs), protein-nucleic acid interactions, and posttranslational modifiers and their targets.

First-generation human binary PPI interactome maps (Rual et al., 2005; Stelzl et al., 2005) have already provided network-based explanations for some genotype-phenotype relationships, but they remain incomplete and of insufficient quality to derive accurate global interpretations (Figure S1A available online). There is a dire need for empirically-controlled (Venkatesan et al., 2009) high-quality proteome-scale interactome reference maps, reminiscent of the high-quality reference genome sequence that revolutionized human genetics.

The challenges are manifold. Even considering only one splice variant per gene, approximately 20,000 protein-coding genes (Kim et al., 2014; Wilhelm et al., 2014) must be handled and ~200 million protein pairs tested to generate a comprehensive binary reference PPI map. Whether such a comprehensive network could ever be mapped by the collective efforts of small-scale studies remains uncertain. Computational predictions of protein interactions can generate information at proteome scale (Zhang et al., 2012) but are inherently limited by biases in currently available knowledge used to infer such interactome models. Should interactome maps be generated for all individual human tissues using biochemical cocomplex association data, or would “context-free” information on direct binary biophysical interaction for all possible PPIs be preferable? To what extent would these approaches be complementary? Even with nearly complete, high-quality reference interactome maps of biophysical interactions, how can the biological relevance of each interaction be evaluated under physiological conditions? Here, we begin to address these questions by generating a proteome-scale map of the human binary interactome and comparing it to alternative network maps.

RESULTS

Vast Uncharted Interactome Zone in Literature

To investigate whether small-scale studies described in the literature are adequate to qualitatively and comprehensively map the human binary PPI network, we assembled all binary pairs identified in such studies and available as of 2013 from seven public databases (Figure S1B, see Extended Experimental Procedures,

Section 1). Out of the 33,000 *literature binary pairs* extracted, two thirds were reported in only a single publication and detected by only a single method (Lit-BS pairs), thus potentially presenting higher rates of curation errors than binary pairs supported by multiple pieces of evidence (Lit-BM pairs; Tables S1A, S1B, and S1C) (Cusick et al., 2009). Testing representative samples from both of these sets using the mammalian protein-protein interaction trap (MAPPIT) (Eyckerman et al., 2001) and yeast two-hybrid (Y2H) (Dreze et al., 2010) assays, we observed that Lit-BS pairs were recovered at rates that were only slightly higher than the randomly selected protein pairs used as negative control (random reference set; RRS) and significantly lower than Lit-BM pairs (Figure 1A and Table S2A; see Extended Experimental Procedures, Section 2). Lit-BS pairs co-occurred in the literature significantly less often than Lit-BM pairs as indicated by STRING literature mining scores (Figure 1A and Figure S1C; see Extended Experimental Procedures, Section 2) (von Mering et al., 2003), suggesting that these pairs were less thoroughly studied. Therefore, use of binary PPI information from public databases should be restricted to interactions with multiple pieces of evidence in the literature. In 2013, this corresponded to 11,045 high-quality protein pairs (Lit-BM-13), more than an order of magnitude below current estimates of the number of PPIs in the full human interactome (Stumpf et al., 2008; Venkatesan et al., 2009).

The relatively low number of high-quality binary literature PPIs may reflect inspection biases inherent to small-scale studies. Some genes such as *RB1* are described in hundreds of publications while most have been mentioned only in a few (e.g., the unannotated *C11orf21* gene). To investigate the effect of such biases on the current coverage of the human interactome network, we organized the interactome search space by ranking proteins according to the number of publications in which they are mentioned (Figure 1B). Interactions between highly studied proteins formed a striking “dense zone” in contrast to a large sparsely populated zone, or “sparse zone,” involving poorly studied proteins. Candidate gene products identified in genome-wide association studies (GWAS) or associated with Mendelian disorders distribute homogeneously across the publication-ranked interactome space (Figure 1B and Figure S1D), demonstrating a need for unbiased systematic PPI mapping to cover this uncharted territory.

A Proteome-wide Binary Interactome Map

Based on literature-curated information, the human interactome appears to be restricted to a narrow dense zone, suggesting that half of the human proteome participates only rarely in the interactome network. Alternatively, the zone that appears sparse in the literature could actually be homogeneously populated by PPIs that have been overlooked due to sociological or experimental biases.

To distinguish between these possibilities and address other fundamental questions outlined above, we generated a new proteome-scale binary interaction map. By acting on all four parameters of our empirically-controlled framework (Venkatesan et al., 2009), we increased the coverage of the human binary interactome with respect to our previous *human interactome* data set obtained by investigating a search space defined by ~7,000 protein-coding genes (“Space I”) and published in 2005 (HI-I-05)

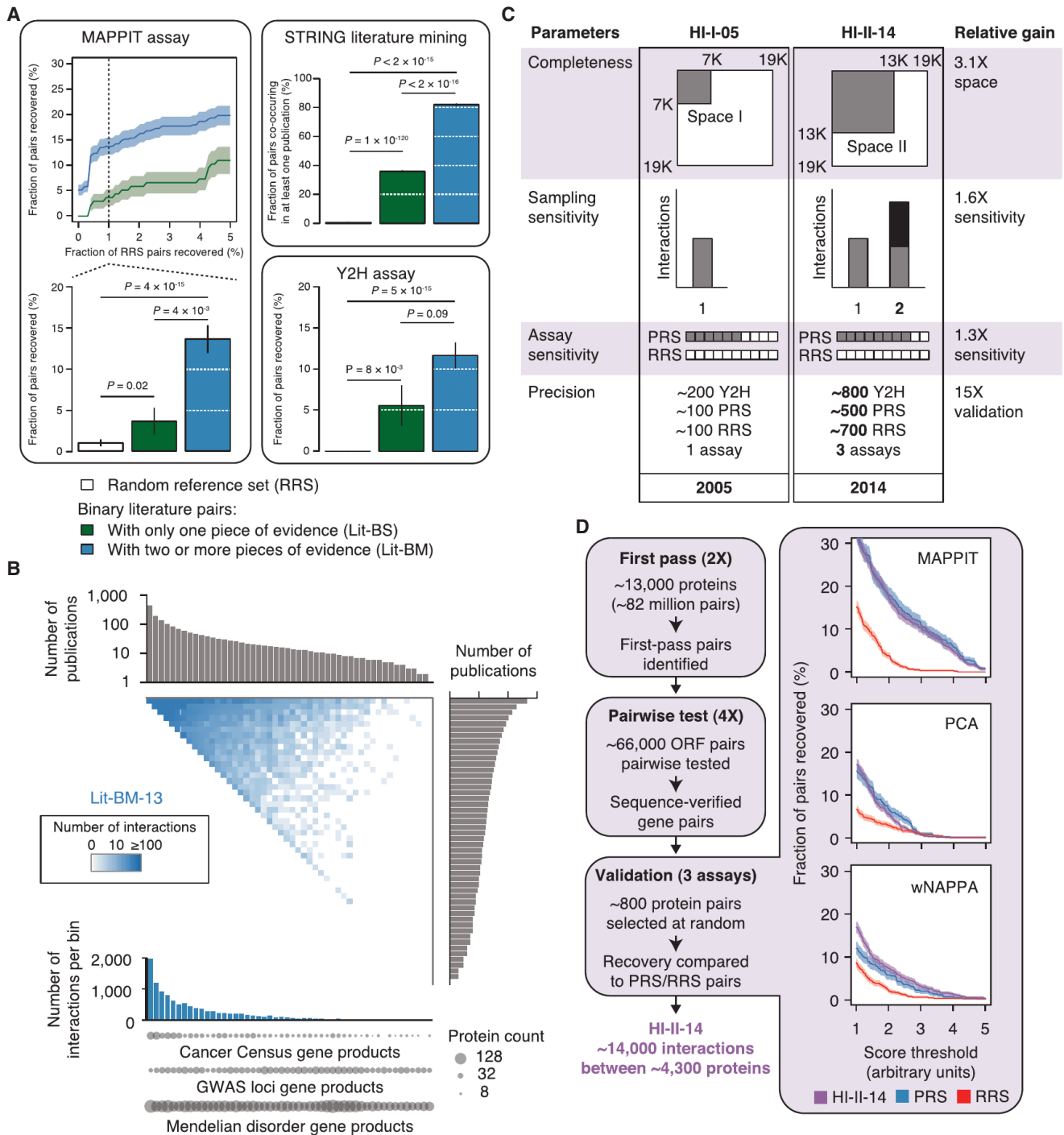


Figure 1. Vast Uncharted Interactome Zone in Literature and Generation of a Systematic Binary Data Set

(A) Validation of binary literature pairs extracted from public databases (Bader et al., 2003; Berman et al., 2000; Chatr-Aryamontri et al., 2013; Kerrien et al., 2012; Licata et al., 2012; Keshava Prasad et al., 2009; Salwinski et al., 2004). Fraction of pairs recovered by MAPPIT at increasing RRS recovery rates (top left) and at 1% RRS recovery rate (bottom left), found to co-occur in the literature as reported in the STRING database (upper right), and recovered by Y2H (lower right). Shading and error bars indicate standard error of the proportion. p values, two-sided Fisher's exact tests. For n values, see Table S6.

(B) Adjacency matrix showing Lit-BM-13 interactions, with proteins in bins of ~350 and ordered by number of publications along both axes. Upper and right histograms show the median number of publications per bin. The color intensity of each square reflects the total number of interactions between proteins for the corresponding bins. Total number of interactions per bin (lower histogram). Number of products from GWAS loci (Hindorf et al., 2009), Mendelian disease (Hamosh et al., 2005), and Sanger Cancer Gene Census (Cancer Census) (Futreal et al., 2004) genes per bin (circles).

(legend continued on next page)

(Rual et al., 2005) (Figures 1C and 1D; see Extended Experimental Procedures, Section 3). A search space consisting of all pairwise combinations of proteins encoded by ~13,000 genes ("Space II"; Table S2B) was systematically probed, representing a 3.1-fold increase with respect to the HI-I-05 search space. To gain in sensitivity, we performed the Y2H assay in different strain backgrounds that showed increased detection of pairs of a positive reference set (PRS) composed of high-quality pairs from the literature without increasing the detection rate of RRS pairs. To increase our sampling, the entire search space was screened twice independently. Pairs identified in this first pass were subsequently tested pairwise in quadruplicate starting from fresh yeast colonies. To ensure reproducibility, only pairs testing positive at least three times out of the four attempts and with confirmed identity were considered interacting pairs, resulting in ~14,000 distinct interacting protein pairs.

We validated these binary interactions using three binary protein interaction assays that rely on different sets of conditions than the Y2H assay: (1) reconstituting a membrane-bound receptor complex in mammalian cells using MAPPIT, (2) *in vitro* using the well-based nucleic acid programmable protein array (wNAPPA) assay (Braun et al., 2009; Ramachandran et al., 2008), and (3) reconstituting a fluorescent protein in Chinese hamster ovary cells using a protein-fragment complementation assay (PCA) (Nyfeler et al., 2005) (see Extended Experimental Procedures, Section 4). The Y2H pairs exhibited validation rates that were statistically indistinguishable from a PRS of ~500 Lit-BM interactions while significantly different from an RRS of ~700 pairs with all three orthogonal assays and over a large range of score thresholds (Figure 1D, Tables S2A and S2C), demonstrating the quality of the entire data set. Using three-dimensional cocrystal structures available for protein complexes in the Protein Data Bank (Berman et al., 2000) and for domain-domain interactions (Stein et al., 2011) (Figure S2 and Tables S2D, S2E, and S2F; see Extended Experimental Procedures, Sections 5 and 6), we also demonstrated that our binary interactions reflect direct biophysical contacts, a conclusion in stark contrast to a previous report suggesting that Y2H interactions are inconsistent with structural data (Edwards et al., 2002). Our results also suggested that Y2H sensitivity correlates with the number of residue-residue contacts and thus presumably with interaction affinity. The corresponding *human interactome* data set covering Space II and reported in 2014 (HI-II-14; Table S2G) is the largest experimentally-determined binary interaction map yet reported, with 13,944 interactions among 4,303 distinct proteins.

Overall Biological Significance

To assess the overall functional relevance of HI-II-14, we combined computational analyses with a large-scale experi-

mental approach. We first measured enrichment for shared Gene Ontology (GO) terms and phenotypic annotations and observed that HI-II-14 shows significant enrichments that are similar to those of Lit-BM-13 (Figures 2A and 2B; see Extended Experimental Procedures, Section 7). Second, we measured how much binary interactions from HI-II-14 reflect membership in larger protein complexes as annotated in CORUM (Ruepp et al., 2010) or reported in a cocomplex association map (Woodsmith and Stelzl, 2014). In both cases, we observed a significant enrichment for binary interactions between protein pairs that belong to a common complex ($p < 0.001$; Figure 2B). Third, we performed a similar analysis using tissue-specific mRNA expression data across the 16 human tissues of the Illumina Human Body Map 2.0 project as well as cellular compartment localization annotations from the GO Slim terms. Again, HI-II-14 was enriched for interactions mediated by protein pairs present in at least one common compartment or cell type (Figures 2C and 2D). Finally, we measured the overlap of HI-II-14 with specific biochemical relationships, as represented by kinase-substrate interactions. Both HI-II-14 and Lit-BM-13 contained significantly more PPIs reflecting known kinase-substrate relationships (Hornbeck et al., 2012) than the corresponding degree-controlled randomized networks (Figure 2E). In addition, HI-II-14 tended to connect tyrosine and serine/threonine kinases (Manning et al., 2002) to proteins with tyrosine or serine/threonine phospho-sites (Hornbeck et al., 2012; Olsen et al., 2010), respectively (Figure S3A), pointing to the corresponding interactions being genuine kinase-substrate interactions. In short, our systematic interactome map, which was generated independently from any pre-existing biological information, reveals functional relationships at levels comparable to those seen for the literature-based interaction map.

To further investigate the overall biological relevance of HI-II-14, we used an experimental approach that compares the impact of mutations associated with human disorders to that of common variants with no reported phenotypic consequences on biophysical interactions (Figure 3). Our rationale is that a set of interactions corresponding to genuine functional relationships should more likely be perturbed by disease-associated mutations than by common variants. The following example will illustrate this concept. Mutations R24C and R24H in CDK4 are clearly associated with melanoma by conferring resistance to CDKN2A inhibition (Wölfel et al., 1995), whereas N41S and S52N mutations are of less clear clinical significance (Zhong et al., 2009) and have remained functionally uncharacterized. HI-II-14 contains five CDK4 interactors: two inhibitors (CDKN2C and CDKN2D), two cyclins (CCND1 and CCND3), and HOOK1, a novel interacting partner and a potential phosphorylation target

(C) Improvements from first-generation to second-generation interactome mapping based on an empirically-controlled framework (Venkatesan et al., 2009). Completeness: fraction of all pairwise protein combinations tested; Assay sensitivity: fraction of all true biophysical interactions that are identifiable by a given assay; Sampling sensitivity: fraction of identifiable interactions that are detected in the experiment; Precision: fraction of reported pairs that are true positives. PRS: positive reference set; RRS: random reference set.

(D) Experimental pipeline for identifying high-quality binary protein-protein interactions (left). ORF: open reading frame. Fraction of HI-II-14, PRS, and RRS pairs (right) recovered by MAPPIT, PCA, and wNAPPA at increasing assay stringency. Shading indicates standard error of the proportion. $p > 0.05$ for all assays when comparing PRS and HI-II-14 at 1% RRS, two-sided Fisher's exact tests. For n values, see Table S6.

See also Figures S1 and S2 and Tables S1 and S2.

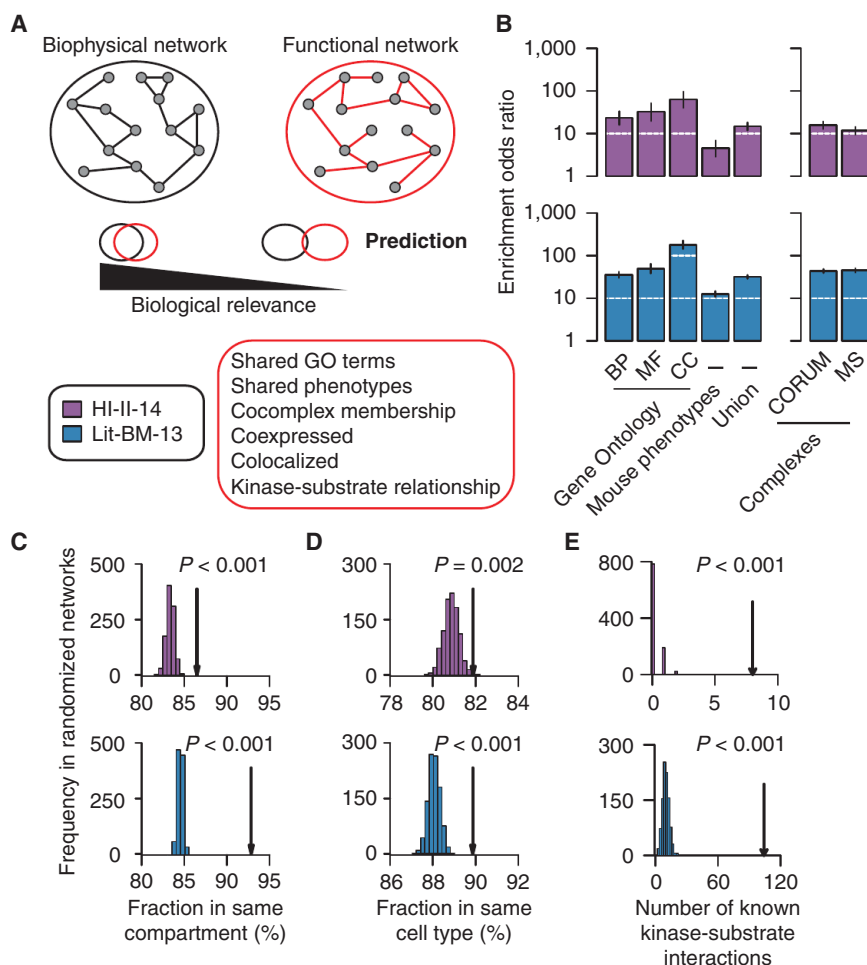


Figure 2. Overall Biological Significance

(A) Schematic of the method to assess biological relevance of binary maps.

(B) Enrichment of binary interactome maps for functional relationships (left) and cocomplex memberships (right). Error bars indicate 95% confidence intervals. BP: Biological process; MF: Molecular function; CC: Cellular component. Mouse phenotypes: Shared phenotypes in mouse models by orthology mapping. MS: Mass-spectrometry-based map. Enrichments: $p \leq 0.05$ for all annotations and maps, two-sided Fisher's exact tests. For n values, see Table S6.

(C) and (D) Fraction of binary interactions between proteins localized in a common cellular compartment and proteins copresent in at least one cell type (arrows) compared to those in 1,000 degree-controlled randomized networks. Empirical p values. For n values, see Table S6.

(E) Number of known kinase-substrate interactions found in binary maps (arrows) compared to those in 1,000 randomized networks. Empirical p values are shown.

See also Figure S3.

of CDK4 (Figure S3B). In agreement with previous reports, the comparative interaction profile shows that R24C and R24H, but not N41S and S52N, specifically perturb CDK4 binding to CDKN2C (Figure 3).

In total, we identified 32 human genes for which: (1) the corresponding gene product is reported to have binary interactors in HI-II-14, (2) germline disease-associated missense mutations have been reported, and (3) common coding missense variants unlikely to be involved in any disease have been identified in the 1000 Genomes Project (1000 Genomes Project Consortium, 2012). To avoid overrepresentation of certain genes, we selected a total of 115 variants, testing up to four disease and four common variants per disease gene for their impact on the ability of the corresponding proteins to interact with known interaction partners (see Extended Experimental Procedures, Section 8). Disease variants were 10-fold more likely to perturb interactions than nondisease variants (Figure 3 and Table S3). Strikingly, more than 55% of the 107 HI-II-14 interactions tested were perturbed by at least one disease-associated variant, and the same trend was observed when considering only mutants with evidence of expression in yeast as indicated by their ability to mediate at least one interaction (Figure S3C). Examples of novel specifically perturbed interactions include AANAT-

BHLHE40 and RAD51D-IKZF1 (Figure 3). In the first case, the A129T mutation in AANAT is known to be associated with a delayed sleeping phase syndrome and specifically perturbs an interaction between AANAT and BHLHE40, the product of a gene reported to function in circadian rhythm regulation (Nakashima et al., 2008). In the second case, the breast-cancer-associated RAD51D E233G mutation perturbs interactions with a number of partners, including the known cancer gene product IKZF1 (Futreal et al., 2004).

Altogether these computational and experimental results provide strong evidence that HI-II-14 pairs correspond to biologically relevant interactions and represent a valuable resource to further our understanding of the human interactome and its perturbations in human disease.

A "Broader" Interactome

Unlike literature-curated interactions, HI-II-14 protein pairs are distributed homogeneously across the interactome space (Figure 4A), indicating that sociological biases, and not fundamental biological properties, underlie the existence of a densely populated zone in the literature. Since 1994, the number of high-quality binary literature PPIs has grown roughly linearly to reach ~11,000 interactions in 2013 (Figure 4B), while systematic data sets are punctuated by a few large-scale releases. Although the sparse territory of the literature map gradually gets populated, interaction density in this zone continues to lag behind that of the dense zone (Figure 4B). In terms of proteome coverage, the expansion rate is faster for systematic maps than for literature maps, especially in the sparse territory (Figure 4C and Figure S4A; see Extended Experimental Procedures,

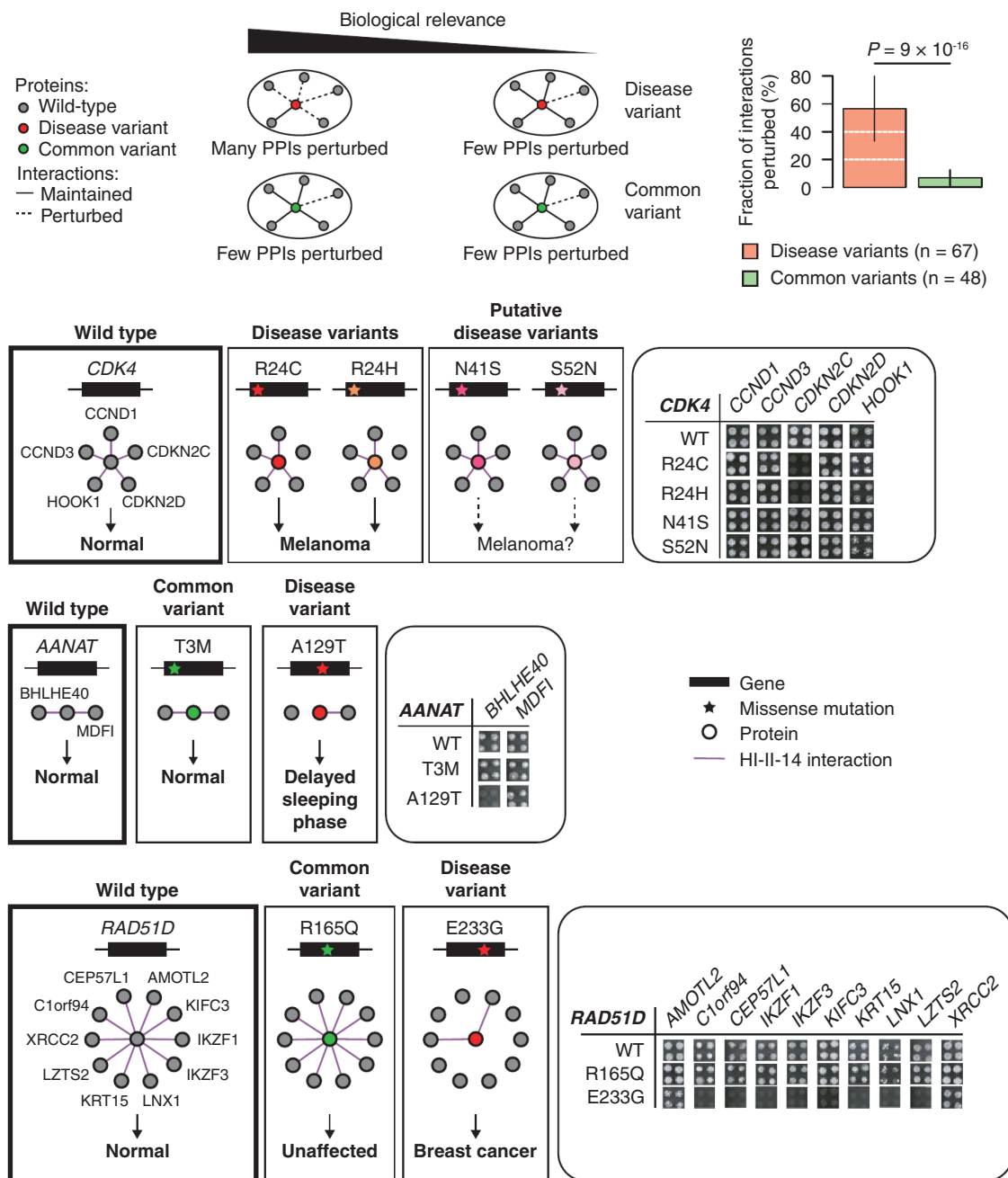


Figure 3. Perturbations of Protein Interactions by Disease and Common Variants

Predicted effect of mutations on PPIs as a function of their biological relevance (top left). Fraction of interactions of the wild-type gene product lost by mutants bearing the disease-associated or common variants (top right, error bars indicate standard error of the proportion). p value, two-sided Fisher's exact test. Comparison of interaction profile of wild-type CDK4, AANAT, and RAD51D to the interaction profile of mutants bearing disease or common variants (bottom). Yeast growth phenotypes on SC-Leu-Trp-His+3AT media in quadruplicate experiments are shown. See also Figure S3 and Table S3.

Section 9). While Lit-BM-13 provides more information in the dense zone, HI-II-14 reveals interactions for more than 2,000 proteins absent from Lit-BM-13. These observations are likely due to a tendency of the literature map to expand from already connected proteins (Figure 4D).

To more deeply explore the heterogeneous coverage of the human interactome, we compared HI-II-14 and Lit-BM-13 to a collection of ~25,000 predicted binary PPIs of high-confidence (PrePPI-HC) (Zhang et al., 2012) and a co-fractionation map of ~14,000 potentially binary interactions (Co-Frac) (Havugimana

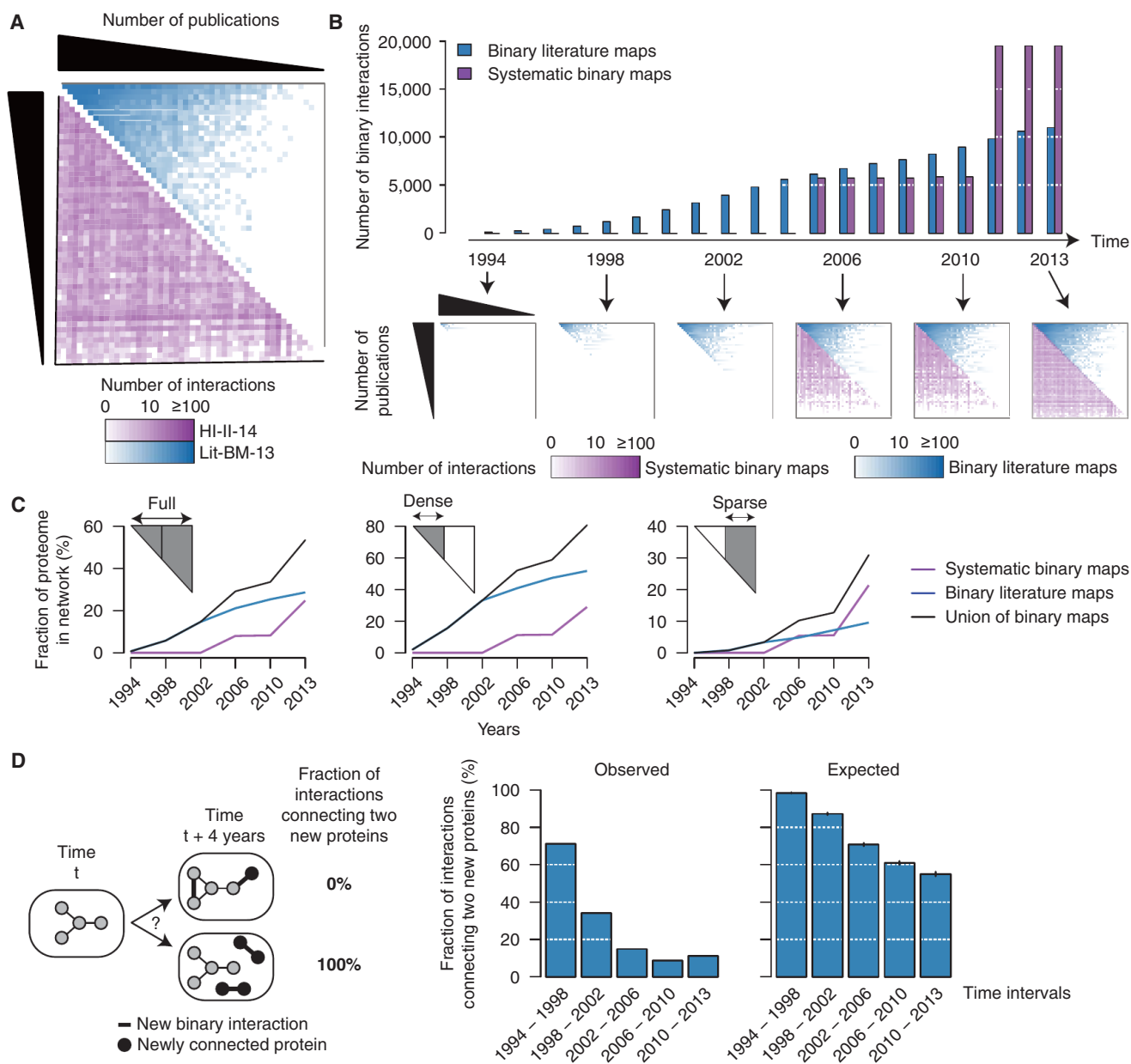


Figure 4. A "Broader" Interactome

(A) Adjacency matrices showing Lit-BM-13 (blue) and HI-II-14 (purple) interactions, with proteins in bins of ~ 350 and ordered by number of publications along both axes. The color intensity of each square reflects the total number of interactions for the corresponding bins.

(B) Total number of binary interactions in literature and systematic interactome maps over the past 20 years (top), with years reflecting either date of public release of systematic binary data sets or date of publication that resulted in inclusion of interactions in Lit-BM-13. Adjacency matrices (bottom) as in Figure 4A.

(C) Fraction of the human proteome present in binary interactome maps at selected time points since 1994, considering the full interactome space (left) or only dense (middle) and sparse (right) zones of Lit-BM-13 with respect to number of publications.

(D) Fraction of new interactions connecting two proteins that were both absent from the map at the previous time point (four years interval; middle) compared to the average in 1,000 randomized networks (right). Error bars indicate standard deviation.

et al., 2012). We tested the extent to which these two data sets contain binary interactions (see Extended Experimental Procedures, Section 10). Representative samples from both Co-Frac and PrePPI-HC were recovered by Y2H at a much lower rate than a sample of Lit-BM-13 and appeared statistically indi-

stinguishable from random pairs (Figure 5A and Table S4A). A *literature non-binary* data set (Lit-NB-13) performed similarly. However, Co-Frac and PrePPI-HC, like Lit-NB-13, were both significantly enriched for functionally relevant relationships. Thus, although these data sets represent potentially valuable

resources, both Co-Frac and PrePPI-HC appear to be more comparable to nonbinary than to binary data sets. Surprisingly, even though PrePPI-HC and Co-Frac systematically surveyed the full human proteome and map different portions of the interactome (Figures S4B), both exhibit a strong tendency to report interactions among well-studied proteins (Figure 5B). This bias is likely due to the integration of functional annotations in the generation of both data sets.

Because coverage might depend on gene expression levels, we also examined interactome maps for expression-related sparse versus dense zones. Co-Frac shows a strong bias toward interactions involving proteins encoded by genes highly expressed in the cell lines used (Figure 5B). This expression-dependent bias is echoed in the literature map, perhaps reflecting a general tendency to study highly expressed proteins. In contrast, both HI-II-14 and PrePPI-HC exhibit a uniform interaction density across the full spectrum of expression levels, likely explained by the standardized expression of proteins tested in Y2H and by the independence of homology-based predictions from expression levels.

We more broadly explored the intrinsic biases that might influence the appearance of sparsely populated zones by examining 21 protein or gene properties, roughly classified as expression-, sequence-, or knowledge-based (Figures 5B and 5C, Tables S4B and S4C; see Extended Experimental Procedures, Section 9). For example, PrePPI-HC is virtually devoid of interactions between proteins lacking Pfam domains, consistent with conserved domains forming the basis of the prediction method. HI-II-14 appears depleted of interactions among proteins containing predicted transmembrane helices, consistent with expected limitations of the Y2H assay (Stagljär and Fields, 2002). Co-Frac is similarly depleted in interactions involving proteins with transmembrane helices, which may result from membrane-bound proteins being filtered out during biochemical fractionations. Compared to HI-II-14, HI-I-05 presented a less homogenous coverage of the space with respect to abundance and knowledge properties, likely reflecting the content of early versions of the hORFeome (Figure S4C). Importantly, no single map appeared unbiased in all 21 examined properties. A combined map presented a slightly increased homogeneity although intrinsic knowledge biases of the three maps using literature-derived evidence were still predominant.

To confirm that HI-II-14 interactions found in the sparse zones of the three other maps are of as high quality as those found in dense zones, we compared MAPPIT validation rates and functional enrichment across these zones for all protein properties examined. MAPPIT validation rates of dense and sparse zone pairs were consistent for nearly all properties (Figures 5D and S4D), indicating that HI-II-14 interactions are of similar biophysical quality throughout the full interactome space. Functional enrichment within the sparse zone was statistically indistinguishable from that of the dense zone (Figures 5D and S4E), demonstrating the functional importance of HI-II-14 biophysical interactions in zones covered sparsely by other types of interactome maps.

Considering all current maps, more than half of the proteome is now known to participate in the interactome network. Our systematic exploration of previously uncharted territories dramatically expands the interactome landscape, suggesting that the

human interactome network is broader in scope than previously observed and that the entire proteome may be represented within a fully mapped interactome.

Interactome Network and Cancer Landscape

Genes associated with the same disease are believed to be preferentially interconnected in interactome networks (Barabási et al., 2011; Vidal et al., 2011). However, in many cases, these observations were made with interactome maps that are composites of diverse evidence, e.g., binary PPIs, cocomplex memberships, and functional associations, a situation further complicated by the uneven quality and sociological biases described above. Using HI-II-14, we revisited this concept for cancer gene products. Our goal was to investigate whether the cancer genomic landscape is limited to the known cancer genes curated in the Sanger Cancer Gene Census (“Cancer Census”) (Futreal et al., 2004), or if, alternatively, it might extend to some of the hundreds of additional candidate genes enriched in somatic mutations uncovered by systematic cancer genome sequencing (“SM genes”) (Chin et al., 2011) and/or identified by functional genomic strategies such as Sleeping Beauty transposon-based screens in mice (“SB genes”) (Copeland and Jenkins, 2010) or global investigations on DNA tumor virus targets (“VT genes”) (Rozenblatt-Rosen et al., 2012).

Given our homogeneous coverage of the space for known (Cancer Census) and candidate (SB, SM, and VT) cancer genes (Figure 6A), we first tested the postulated central role of cancer gene products in biological networks (Barabási et al., 2011) and verified that both sets tend to have more interactions and to be more central in the systematic map than proteins not associated with cancer (Figure 6B). We then examined the interconnectivity of known cancer proteins and showed that Cancer Census gene products interact with each other more frequently than expected by chance, a trend not apparent in HI-I-05 (Figure 6C). We sought to use this topological property as the basis for novel cancer gene discovery in the large lists of cancer candidates from genomic and functional genomic screens.

We examined whether products of candidate cancer genes identified by GWAS (Table S5A) tend to be connected to Cancer Census proteins, and observed significant connectivity in all four maps (Figure S5A; see Extended Experimental Procedures, Section 11). When loci containing a known cancer gene were excluded, only HI-II-14 showed such connectivity, supporting its unique value to identify cancer candidate genes beyond those already well demonstrated (Figures 7A and S5A). In further support of their association with cancer, genes in cancer GWAS loci prioritized by “guilt-by-association” in HI-II-14 tend to correspond to cancer candidates from systematic cancer studies (Figures 7B and 7C). These results suggest that cancer-associated proteins tend to form subnetworks perturbed in tumorigenesis, and that HI-II-14 provides new context to prioritize cancer genes from genome-wide studies.

The following example illustrates the power of our combined approach. C-terminal Binding Protein 2 (CTBP2) is encoded at a locus associated with prostate cancer susceptibility (Thomas et al., 2008) and belongs to both SB and VT gene lists (Mann et al., 2012; Rozenblatt-Rosen et al., 2012). Two Cancer Census genes, *IKZF1* and *FLI1*, encode interacting partners of

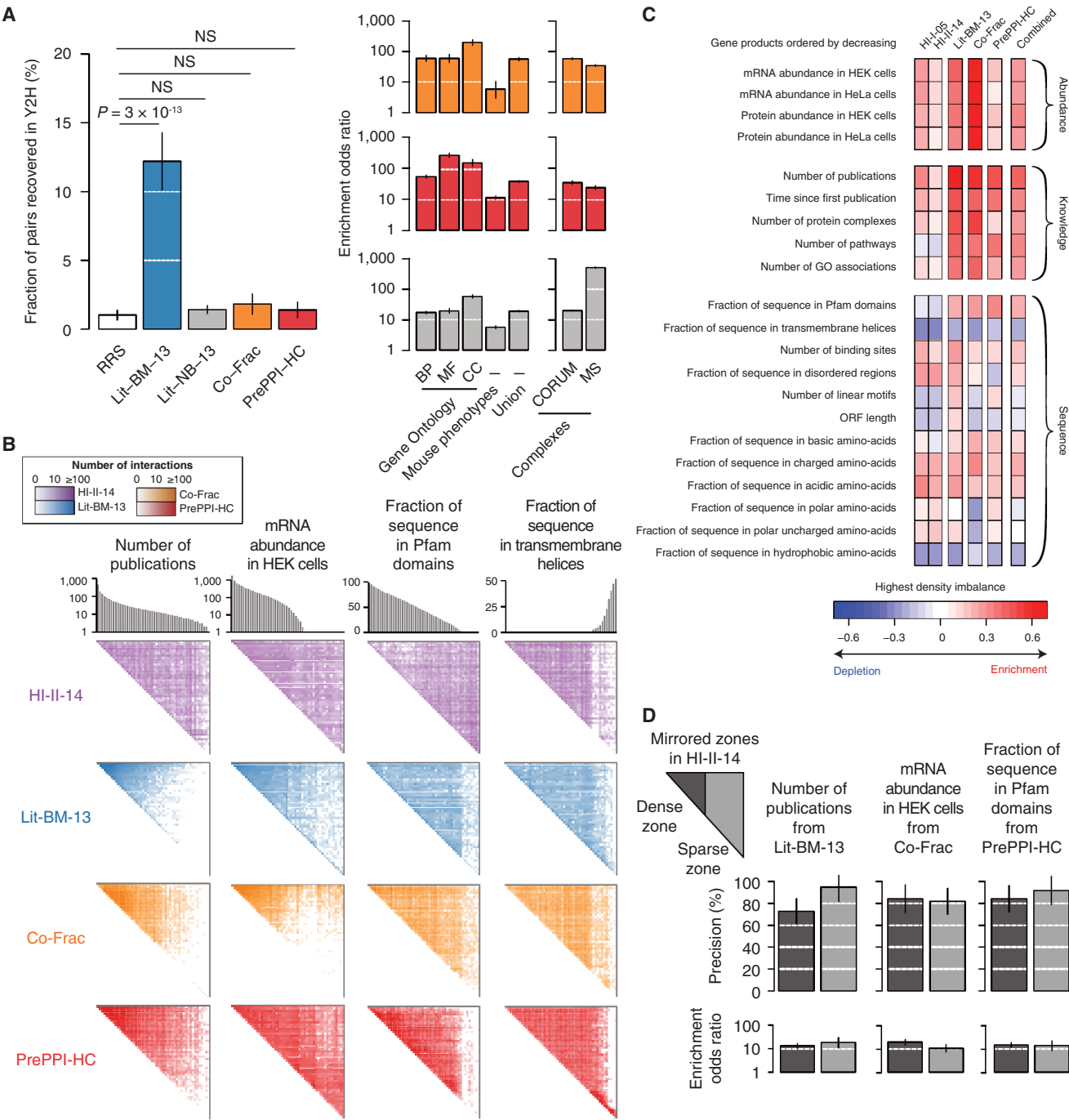


Figure 5. Comparison of Interaction Mapping Approaches

(A) Evaluation of the quality of Co-Frac (orange), PrePPI-HC (red), and pairs from small-scale experiments in the literature with no binary evidence (Lit-NB-13, grey). Fraction of pairs recovered by Y2H as compared to pairs from Lit-BM-13 and pairs of randomly selected proteins (RRS) (left). Error bars indicate standard error of the proportion. Enrichment in functional interactions and cocomplex memberships (right). Legend as in Figure 2B. For n values, see Table S6.

(B) Adjacency matrices for HI-II-14, Lit-BM-13, Co-Frac, and PrePPI-HC maps, with proteins per bins of ~ 350 and ordered by number of publications, mRNA abundance in HEK cells, fraction of protein sequence covered by Pfam domains, or fraction of protein sequence in transmembrane helices. Figure legend as in Figure 1B.

(C) Highest interaction density imbalances (observed minus expected) in the four maps, the union of all four maps, and our previous binary map (HI-II-05) for 21 protein properties.

(legend continued on next page)

CTBP2 in HI-II-14. These are transcription factors with tumor suppressor (Payne and Dovat, 2011) and proto-oncogene (Kornblau et al., 2011) roles, respectively, in lymphoid tumors. Given its interactions with IKZF1 and FLI1, we investigated the potential role of CTBP2 in lymphoid tumorigenesis. In the Cancer Cell Line Encyclopedia (Barretina et al., 2012), *FLI1* was significantly more often amplified in lymphoid than in other cell lines (Figure 7D), consistent with its proposed proto-oncogenic role in these tumors. In contrast, both *CTBP2* and *IKZF1*, but not *CTBP1*, were deleted significantly more often in lymphoid cancer cell lines. Notably, deletion of *CTBP2* or *IKZF1* and amplification of *FLI1* were mostly nonoverlapping in the different cell lines, suggesting that either event may be sufficient to affect tumorigenesis (Figure S5B). Altogether, these results suggest a role for CTBP2 in suppressing lymphoid tumors by direct repression of FLI1 function, potentially involving IKZF1.

Finally, we assessed how HI-II-14 interactions can be integrated with genomic and functional genomic data sets. Going beyond the “guilt-by-profiling” concept, we also used these gene sets in “guilt-by-association” predictions in a combined model (Figure S6A), which leads to substantially improved cancer gene rankings over those found using either predictive strategy alone (Figures 7E, S6B, and S6C and Table S5B; see Extended Experimental Procedures, Section 12). In contrast, a similar analysis using HI-I-05 interactions showed that its limited size prevented inclusion of any guilt-by-association terms (Figure S6D). Genes significantly mutated in cancer patients from recent TCGA pan-cancer mutation screens (Table S5C) (Lawrence et al., 2014) were enriched among highly ranked predictions from the combined model ($p = 6 \times 10^{-3}$, one-sided Wilcoxon rank test), supporting the validity of our integrated cancer gene predictions. Our top-ranked prediction was the cyclin-dependent kinase 4 (CDK4), a well-known cancer gene product. Four other genes from the Cancer Census list appeared among the top 25 ranked genes. Strikingly, STAT3, which ranked third, was added to the Cancer Census after our training set was established, highlighting the ability of this approach to identify novel cancer gene products.

To characterize the biological processes in which the candidate cancer genes predicted by the combined model are likely to be involved, we identified binary interactions linking them to each other or to Cancer Census proteins in the 12 “pathways of cancer” relevant to cancer development and progression (Table S5D) (Vogelstein et al., 2013). Of our top 100 candidates, 60 mapped to at least one cancer pathway (Figures 7F and S7), twice as many as would be expected from predictions using either the guilt-by-profiling or guilt-by-association approach alone. We propose that many novel cancer candidates can be annotated to specific processes based on their interactions with Cancer Census gene products and known participation in cellular pathways. For example, the candidate protein ID3, a DNA-binding inhibitor, interacts with the two Cancer Census transcription

factors TCF12 and TCF3, suggesting a role for ID3 in the regulation of transcription by inhibiting binding of specific transcription factors to DNA (Loveys et al., 1996; Richter et al., 2012). CTBP2, which we identified as a potential suppressor in lymphoid tumors, represents another example (Figures 5E and S7).

In summary, the increased and uniform coverage of HI-II-14 demonstrates that known and candidate cancer gene products are highly connected in the interactome network, which in turn provides unbiased evidence for an expanded functional cancer landscape.

DISCUSSION

By systematically screening half of the interactome space with minimal inspection bias, we more than doubled the number of high-quality binary PPIs available from the literature. Covering zones of the human interactome landscape that have been weakly charted by other approaches, our systematic binary map provides deeper functional context to thousands of proteins, as demonstrated for candidates identified in unbiased cancer genomic screens. Systematic binary mapping therefore stands as a powerful approach to “connect the dots” of the genomic revolution.

Combining high-quality binary pairs from the literature with systematic binary maps, 30,000 high-confidence interactions are now available. It is likely that a large proportion of the human interactome can soon be mapped by taking advantage of the emergence of reference proteome maps (Kim et al., 2014; Wilhelm et al., 2014), a combination of nearly complete clone collections (Yang et al., 2011), rapid improvements in Y2H assay sensitivity, and emerging interaction-mapping technologies that drastically reduce cost (Caufield et al., 2012; Stagljar and Fields, 2002; Yu et al., 2011).

Reference binary interactome maps of increased coverage and quality will be required to interpret condition-specific interactions and to characterize the effects of splicing and genetic variation on interactions (Zhong et al., 2009). While protein-protein interactions represent an important class of interactions between macromolecules, future efforts integrating this information with protein-DNA, protein-RNA, RNA-RNA or protein-metabolite interactions will provide a unified view of the molecular interactions governing cell behavior. Just as a reference genome enabled detailed maps of human genetic variation (1000 Genomes Project Consortium, 2012), completion of a reference interactome network map will enable deeper insight into genotype-phenotype relationships in human.

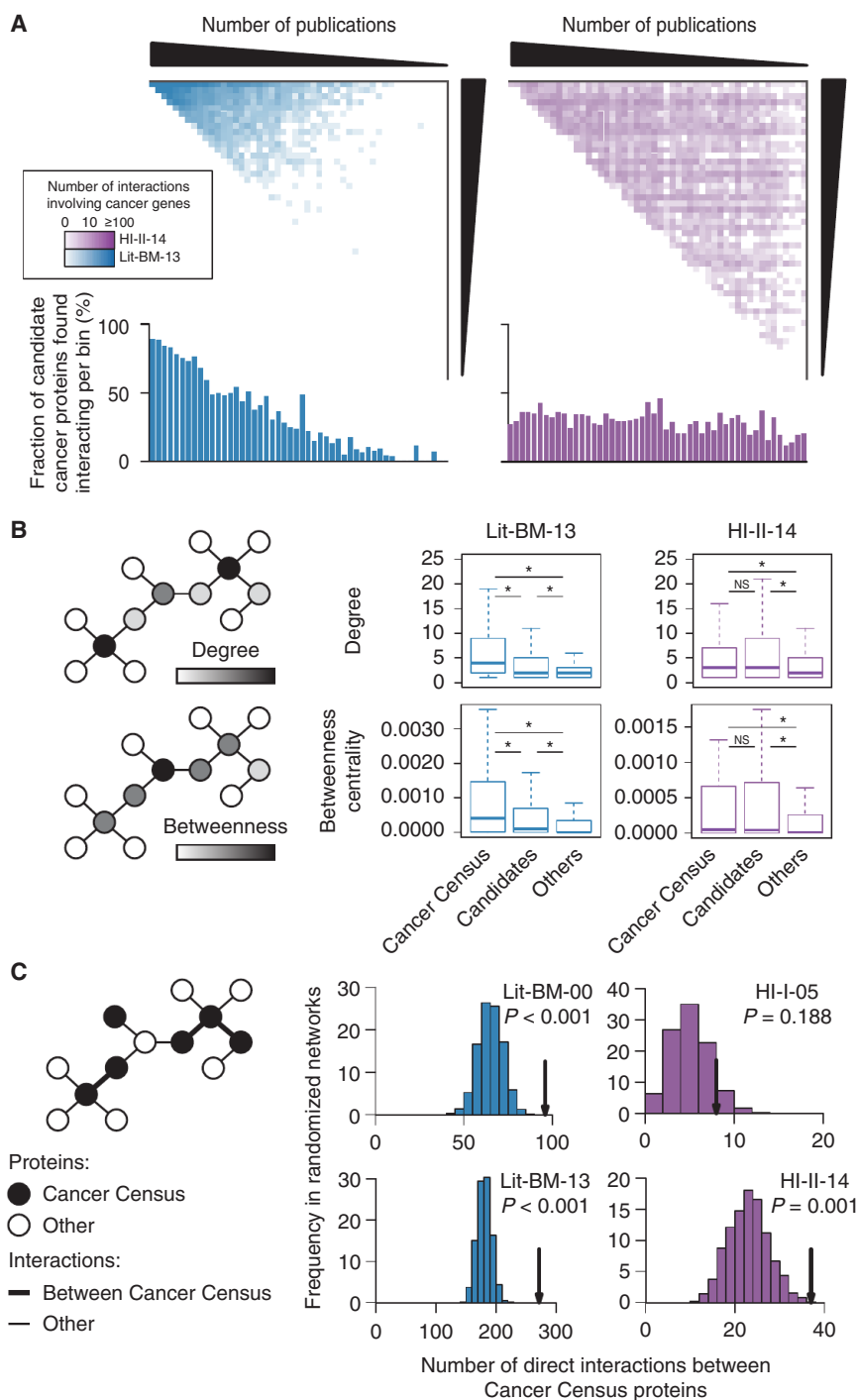
EXPERIMENTAL PROCEDURES

Extraction of the Literature-Based Data Sets

Human PPIs annotated with tractable publication records were extracted from seven databases through August 2013. Large-scale systematic data sets and

(D) Precision at 1% RRS recovery in the MAPFIT assay (top, error bars indicate standard error of the proportion) and functional enrichment (bottom, union of Gene Ontology and mouse phenotypes based annotations, error bars indicate 95% confidence intervals) of HI-II-14 pairs found in dense and sparse zones mirrored from Lit-BM-13, Co-Frac, and PrePPI-HC. $p > 0.05$ for all pairwise comparisons of dense and sparse zones, two-sided Fisher's exact tests. For n values, see Table S6.

See also Figure S4 and Table S4.



pairs involving products of *UBC*, *SUMO1*, *SUMO2*, *SUMO3*, *SUMO4*, or *NEDD8*, were excluded. The remaining pairs were divided into those having no pieces of binary evidence (Lit-NB) and those with at least one piece of binary evidence based on PSI-MI experimental method codes. Binary pairs were divided between pairs with one and with two or more pieces of evidence (Lit-BS and Lit-BM, respectively). For benchmark experiments in Y2H, MAPPIIT, PCA, and wNAPPA, equivalent data sets were extracted similarly in December 2010.

given threshold was calculated as the fraction of PPIs *observed* in the first region minus the fraction of PPIs *expected* assuming a uniform distribution in the space. Dense and sparse zones were defined by identifying the threshold for which the deviation from expectation is maximal.

Measure of Functional Enrichment

For each pairwise comparison, PPI and functional maps were trimmed to pairs where both proteins were present in both maps and restricted to Space II to

Figure 6. Network Properties of Cancer Gene Products

(A) Adjacency matrices for Lit-BM-13 and HI-II-14 only showing interactions involving the product of a Cancer Census (Futreal et al., 2004) or of a candidate cancer gene. Figure legend as in Figure 1B. Lower histograms show for each bin, the fraction of cancer candidates having at least one interaction.

(B) Distribution of the number of interactions (degree) and normalized number of shortest paths between proteins (betweenness centrality) for products of Cancer Census and of candidate cancer genes in Lit-BM-13 and in HI-II-14 maps as compared to other proteins (right; * for $p < 0.05$, NS for $p > 0.05$, two-sided Wilcoxon rank sum tests). For n values, see Table S6.

(C) Number of interactions between products of Cancer Census genes (arrows) in HI-I-05, HI-II-14, Lit-BM as of 2000 (Lit-BM-00) and as of 2013 (Lit-BM-13), as compared to 1,000 degree-controlled randomized networks. Empirical p values. For n values, see Table S6.

Generation of the Binary Protein-Protein Interaction Map

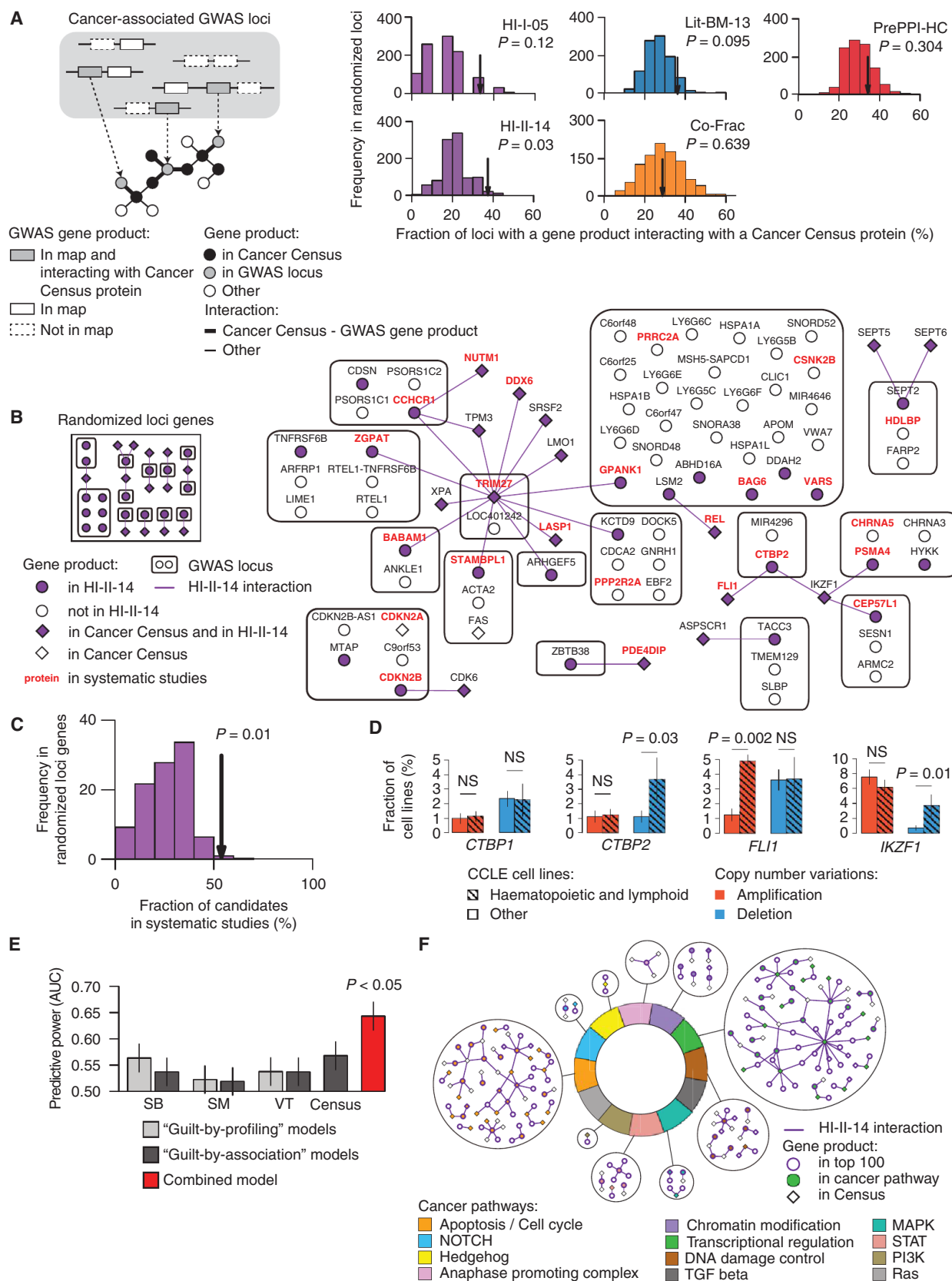
HI-II-14 was generated by screening all pairwise combinations of 15,517 ORFs from hORFeome v5.1 (Space II) as described previously (Dreze et al., 2010). ORFs encoding first pass pairs were identified either by Sanger sequencing or by Stitch-seq (Yu et al., 2011). HI-II-14 was validated by comparing a subset of 809 interactions to a positive and a random reference set of 460 and 698 protein pairs, respectively, using MAPPIIT, PCA, and wNAPPA assays.

Interaction Perturbation by Missense Mutations

Disease variants were obtained from the Human Gene Mutation Database (HGMD 2009 V2) (Stenson et al., 2014) and common variants were derived from the 1000 Genomes Project (1000 Genomes Project Consortium, 2012). Only variants with a minor allele frequency above 1% were considered common. All successfully cloned disease and common variants were systematically tested for interaction with all interactors of their wild-type counterpart.

Interaction Density Imbalance

For each protein property, we ranked all proteins and, for any property threshold, partitioned the interactome space into a first region containing pairs of proteins both above (or below) the threshold, and a second region containing all remaining pairs. Interaction density imbalance of a given map for a



(legend on next page)

allow comparison between PPI maps. Functional enrichment odds ratios were calculated using Fisher's exact tests.

GWAS Analysis

307 distinct cancer-associated SNPs were identified from 75 GWAS publications covering 10 types of cancer and 142 distinct loci were identified at a linkage disequilibrium threshold of 0.9. For each map, we calculated the number of loci encoding an interactor of a Cancer Census protein over the number of loci encoding a protein in the PPI map. To assess significance, we measured the corresponding fraction when randomly selecting for each locus the same number of proteins than genes with products in the PPI map.

Cancer Association Scoring System

For each gene, seven features were measured. Three features represent membership in the SB, SM, and VT lists of candidate cancer genes ("guilt-by-profiling" features). The four other features represent its number of interactors in HI-II-14 that are present in these three lists and in the Cancer Census list, normalized by the expected numbers in degree-controlled randomized networks ("guilt-by-association" features). We measured the ability of each feature to prioritize known Cancer Census genes with separate logistic regression models. We combined all seven features in a forward stepwise logistic regression model using the Akaike information criterion to determine the stepwise halting. The final set of features selected was: the SB, SM, and VT guilt-by-profiling and the Cancer Census and SB guilt-by-association features. "Receiver Operating Characteristic" curves were obtained by measuring at decreasing score threshold the fraction of known Cancer Census genes recovered and the corresponding fraction of proteins predicted as candidate cancer genes.

Data Sets

For reference data sets used in this study, see Extended Experimental Procedures, Section 13. All high-quality binary PPIs described in this paper can be accessed at: http://interactome.dfci.harvard.edu/H_sapiens/.

SUPPLEMENTAL INFORMATION

Supplemental Information includes Extended Experimental Procedures, seven figures, and six tables and can be found with this article online at <http://dx.doi.org/10.1016/j.cell.2014.10.050>.

AUTHOR CONTRIBUTIONS

Computational analyses were performed by T.R., M.T., B.C., S.J.P., C. Fontanillo, R.M., A.K., and S.D.G. with help from A.-R.C., J.C.-H., C. Fan, E.F., M.J., S.K., G.N.L., K.L., J.M., A. Sharma, and Y.S. Experiments were performed by Q.Z., N.S., S.Y., I.L., X.Y., and L.G. with help from D.B., B.E.B., P.B., M.B.,

M.P.B., D.C.-Z., R.C., E.D., M.D., A.D., F.G., B.J.G., M.F.H., R.K., A.M., R.R.M., A.P., M.M.P., X.R., J.R., P.R., V.R., E.R., J.M.S., A. Scholz, A.A.S., K.S., S.T., A.O.T., S.A.T., J.-C.T., K.V., and J.W. Structural analyses were done by T.R., M.T., and R.M. Extraction of the literature data sets was performed by C. Fontanillo, A.K., Ch.F., M.E.C., and T.H. MAPPIT validation was done by I.L. The adjacency matrix interactome representation was developed by S.J.P. with M.T. Functional enrichment analysis was done by T.R., M.T., and B.C. Interaction perturbation experiments were performed by N.S. and S.Y. with Q.Z. Interactome and proteome coverage analyses were done by T.R. and B.C. Comparison of alternative maps was done by T.R., M.T., B.C., and S.J.P. The density imbalance measure was conceived by M.T. Topological analyses were done by T.R., M.T., B.C., S.J.P., and S.D.G. Cancer-related analyses were done by T.R., M.T., and B.C. The cancer association scoring system was done by M.T. CTBP2 and cancer landscape analyses were done by T.R. Interactome mapping was supervised by D.E.H. and M.V. Principal investigators overseeing primary data management, structural biology, literature recuration and reference set construction, MAPPIT validation, and other computational analyses were T.H., P.A., J.D.L.R., J.T., and F.P.R., respectively. B.C., Y.X., A.-L.B., L.M.I., P.A., J.D.L.R., J.T., M.A.C., D.E.H., T.H., F.P.R., and M.V. designed and/or advised the overall research effort. T.R., M.T., B.C., Q.Z., M.E.C., J.D.L.R., M.A.C., D.E.H., T.H., F.P.R., and M.V. wrote the manuscript with contributions from other coauthors.

ACKNOWLEDGMENTS

The authors wish to acknowledge past and present members of the Center for Cancer Systems Biology (CCSB) and particularly H. Yu for helpful discussions. This work was supported primarily by NHGRI grant R01/U01HG001715 awarded to M.V., D.E.H., F.P.R., and J.T. and in part by the following grants and agencies: NHGRI P50HG004233 to M.V., F.P.R., and A.-L.B.; NHLBI U01HL098166 subaward to M.V.; NHLBI U01HL108630 subaward to A.-L.B.; NCI U54CA112962 subaward to M.V.; NCI R33CA132073 to M.V.; NIH RC4HG006066 to M.V., D.E.H., and T.H.; NICHD ARRA R01HD065288, R21MH104766, and R01MH105524 to L.M.I.; NIMH R01MH091350 to L.M.I. and T.H.; NSF CCF-1219007 and NSERC RGPIN-2014-03892 to Y.X.; Canada Excellence Research Chair, Krembil Foundation, Ontario Research Fund-Research Excellence Award, Avon Foundation, grant CS107A09 from Junta de Castilla y Leon (Valladolid, Spain), grant PI12/00624 from Ministerio de Economia y Competitividad (AES 2012, ISCIII, Madrid, Spain), and grant i-Link0398 from Consejo Superior de Investigaciones Científicas (CSIC, Madrid, Spain) to J.D.L.R.; Spanish Ministerio de Ciencia e Innovación (BIO2010-22073) and the European Commission through the FP7 project SyStemAge grant agreement n:306240 to P.A.; Group-ID Multidisciplinary Research Partnerships of Ghent University, grant FWO-V G.0864.10 from the Fund for Scientific Research-Flanders and ERC Advanced Grant N° 340941 to J.T.; EMBO long-term fellowship to A.K.; Institute Sponsored Research funds from the

Figure 7. Interactome Network and Cancer Landscape

(A) Fraction of cancer-related GWAS loci containing at least one gene encoding a protein that interacts with the product of a Cancer Census gene in HI-I-05, HI-II-14, Lit-BM-13, Co-Frac, and PrePPI-HC (arrows) as compared to randomly selected loci genes. GWAS loci already containing a Cancer Census gene are excluded. Empirical p values. For n values, see Table S6.

(B) Network representing products of genes in cancer-associated GWAS loci and their interactions with Cancer Census proteins in HI-II-14 (right), and a representative example of the network obtained for randomized loci genes (left).

(C) Fraction of GWAS loci gene products interacting with a Cancer Census protein also identified in systematic genomic and functional genomic studies (arrow) as compared to the fraction obtained for randomized loci genes (bottom right). Empirical p value.

(D) *CTBP2* and *IKZF1* are deleted in significantly more hematopoietic and lymphoid cancer cell lines than in other cancer cell lines. CCLE, Cancer Cell Line Encyclopedia. Each barplot compares the fraction of cell lines from the 163 hematopoietic and lymphoid (hatched bars) or 717 other (empty bars) cell types where *CTBP1*, *CTBP2*, *FLI1*, or *IKZF1* were found amplified (red) or deleted (blue). Error bars indicate standard error of the proportion. p values, two-sided Fisher's exact tests (NS for $p > 0.05$).

(E) Predictive power of guilt-by-profiling and guilt-by-association models compared to the combined model (Figure S6; see Extended Experimental Procedures, Section 11). AUC: Area under the curve in Figure S6C. Error bars indicate standard error of the proportion. p value, two-sided Wilcoxon rank sum test. SB, Sleeping Beauty transposon-based mouse cancer screen; SM, Somatic mutation screen in cancer tissues; VT, Virus targets.

(F) Binary interactions from HI-II-14 involving the top candidates and Cancer Census gene products in the twelve pathways associated to cancer development and progression.

See also Figures S5, S6, and S7 and Table S5.

Dana-Farber Cancer Institute Strategic Initiative to M.V. I.L. is a postdoctoral fellow with the FWO-V. M.V. is a "Chercheur Qualifié Honoraire" from the Fonds de la Recherche Scientifique (FRS-FNRS, Wallonia-Brussels Federation, Belgium). Since performing the work described, C. Fontanillo has become an employee of Celgene Research SL, part of the Celgene Corporation.

Received: September 17, 2014

Revised: October 21, 2014

Accepted: October 30, 2014

Published: November 20, 2014

REFERENCES

- 1000 Genomes Project Consortium, Abecasis, G.R., Auton, A., Brooks, L.D., DePristo, M.A., Durbin, R.M., Handsaker, R.E., Kang, H.M., Marth, G.T., and McVean, G.A. (2012). An integrated map of genetic variation from 1,092 human genomes. *Nature* 491, 56–65.
- Bader, G.D., Betel, D., and Hogue, C.W. (2003). BIND: the Biomolecular Interaction Network Database. *Nucleic Acids Res.* 31, 248–250.
- Barabási, A.-L., Gulbahce, N., and Loscalzo, J. (2011). Network medicine: a network-based approach to human disease. *Nat. Rev. Genet.* 12, 56–68.
- Barretina, J., Caponigro, G., Stransky, N., Venkatesan, K., Margolin, A.A., Kim, S., Wilson, C.J., Lehár, J., Kryukov, G.V., Sonkin, D., et al. (2012). The Cancer Cell Line Encyclopedia enables predictive modelling of anticancer drug sensitivity. *Nature* 483, 603–607.
- Berman, H.M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T.N., Weissig, H., Shindyalov, I.N., and Bourne, P.E. (2000). The Protein Data Bank. *Nucleic Acids Res.* 28, 235–242.
- Braun, P., Taşan, M., Dreze, M., Barrios-Rodiles, M., Lemmens, I., Yu, H., Sahalie, J.M., Murray, R.R., Roncari, L., de Smet, A.S., et al. (2009). An experimentally derived confidence score for binary protein-protein interactions. *Nat. Methods* 6, 91–97.
- Caufield, J.H., Sakhawalkar, N., and Uetz, P. (2012). A comparison and optimization of yeast two-hybrid systems. *Methods* 58, 317–324.
- Chatr-Aryamontri, A., Breitkreutz, B.J., Heinicke, S., Boucher, L., Winter, A., Stark, C., Nixon, J., Ramage, L., Kolas, N., O'Donnell, L., et al. (2013). The BioGRID interaction database: 2013 update. *Nucleic Acids Res.* 41 (Database issue), D816–D823.
- Chin, L., Hahn, W.C., Getz, G., and Meyerson, M. (2011). Making sense of cancer genomic data. *Genes Dev.* 25, 534–555.
- Copeland, N.G., and Jenkins, N.A. (2010). Harnessing transposons for cancer gene discovery. *Nat. Rev. Cancer* 10, 696–706.
- Cusick, M.E., Yu, H., Smolyar, A., Venkatesan, K., Carvunis, A.-R., Simonis, N., Rual, J.F., Borick, H., Braun, P., Dreze, M., et al. (2009). Literature-curated protein interaction datasets. *Nat. Methods* 6, 39–46.
- Dreze, M., Monachello, D., Lurin, C., Cusick, M.E., Hill, D.E., Vidal, M., and Braun, P. (2010). High-quality binary interactome mapping. *Methods Enzymol.* 470, 281–315.
- Edwards, A.M., Kus, B., Jansen, R., Greenbaum, D., Greenblatt, J., and Gerstein, M. (2002). Bridging structural biology and genomics: assessing protein interaction data with known complexes. *Trends Genet.* 18, 529–536.
- Eyckerman, S., Verhee, A., Van der Heyden, J., Lemmens, I., Ostade, X.V., Vandekerckhove, J., and Tavernier, J. (2001). Design and application of a cytokine-receptor-based interaction trap. *Nat. Cell Biol.* 3, 1114–1119.
- Futreal, P.A., Coin, L., Marshall, M., Down, T., Hubbard, T., Wooster, R., Rahman, N., and Stratton, M.R. (2004). A census of human cancer genes. *Nat. Rev. Cancer* 4, 177–183.
- Hamosh, A., Scott, A.F., Amberger, J.S., Bocchini, C.A., and McKusick, V.A. (2005). Online Mendelian Inheritance in Man (OMIM), a knowledgebase of human genes and genetic disorders. *Nucleic Acids Res.* 33 (Database issue), D514–D517.
- Havugimana, P.C., Hart, G.T., Nepusz, T., Yang, H., Turinsky, A.L., Li, Z., Wang, P.I., Boutz, D.R., Fong, V., Phanse, S., et al. (2012). A census of human soluble protein complexes. *Cell* 150, 1068–1081.
- Hindorff, L.A., Sethupathy, P., Junkins, H.A., Ramos, E.M., Mehta, J.P., Collins, F.S., and Manolio, T.A. (2009). Potential etiologic and functional implications of genome-wide association loci for human diseases and traits. *Proc. Natl. Acad. Sci. USA* 106, 9362–9367.
- Hornbeck, P.V., Kornhauser, J.M., Tkachev, S., Zhang, B., Skrzypek, E., Murray, B., Latham, V., and Sullivan, M. (2012). PhosphoSitePlus: a comprehensive resource for investigating the structure and function of experimentally determined post-translational modifications in man and mouse. *Nucleic Acids Res.* 40 (Database issue), D261–D270.
- International Human Genome Sequencing Consortium (2004). Finishing the euchromatic sequence of the human genome. *Nature* 431, 931–945.
- Kerrien, S., Aranda, B., Breuza, L., Bridge, A., Broackes-Carter, F., Chen, C., Duesbury, M., Dumousseau, M., Feuermann, M., Hinz, U., et al. (2012). The IntAct molecular interaction database in 2012. *Nucleic Acids Res.* 40 (Database issue), D841–D846.
- Keshava Prasad, T.S., Goel, R., Kandasamy, K., Keerthikumar, S., Kumar, S., Mathivanan, S., Telikicherla, D., Raju, R., Shafreen, B., Venugopal, A., et al. (2009). Human Protein Reference Database—2009 update. *Nucleic Acids Res.* 37 (Database issue), D767–D772.
- Kim, M.S., Pinto, S.M., Getnet, D., Nirujogi, R.S., Manda, S.S., Chaerkady, R., Madugundu, A.K., Kelkar, D.S., Isserlin, R., Jain, S., et al. (2014). A draft map of the human proteome. *Nature* 509, 575–581.
- Kornblau, S.M., Qiu, Y.H., Zhang, N., Singh, N., Faderl, S., Ferrajoli, A., York, H., Qutub, A.A., Coombes, K.R., and Watson, D.K. (2011). Abnormal expression of FLI1 protein is an adverse prognostic factor in acute myeloid leukemia. *Blood* 118, 5604–5612.
- Lawrence, M.S., Stojanov, P., Mermel, C.H., Robinson, J.T., Garraway, L.A., Golub, T.R., Meyerson, M., Gabriel, S.B., Lander, E.S., and Getz, G. (2014). Discovery and saturation analysis of cancer genes across 21 tumour types. *Nature* 505, 495–501.
- Licata, L., Briganti, L., Peluso, D., Perfetto, L., Iannuccelli, M., Galeota, E., Sacco, F., Palma, A., Nardoza, A.P., Santonico, E., et al. (2012). MINT, the molecular interaction database: 2012 update. *Nucleic Acids Res.* 40 (Database issue), D857–D861.
- Loveys, D.A., Streiff, M.B., and Kato, G.J. (1996). E2A basic-helix-loop-helix transcription factors are negatively regulated by serum growth factors and by the Id3 protein. *Nucleic Acids Res.* 24, 2813–2820.
- Mann, K.M., Ward, J.M., Yew, C.C., Kovochich, A., Dawson, D.W., Black, M.A., Brett, B.T., Sheetz, T.E., Dupuy, A.J., Chang, D.K., et al.; Australian Pancreatic Cancer Genome Initiative (2012). Sleeping Beauty mutagenesis reveals cooperating mutations and pathways in pancreatic adenocarcinoma. *Proc. Natl. Acad. Sci. USA* 109, 5934–5941.
- Manning, G., Whyte, D.B., Martinez, R., Hunter, T., and Sudarsanam, S. (2002). The protein kinase complement of the human genome. *Science* 298, 1912–1934.
- Nakashima, A., Kawamoto, T., Honda, K.K., Ueshima, T., Noshiro, M., Iwata, T., Fujimoto, K., Kubo, H., Honma, S., Yorioka, N., et al. (2008). DEC1 modulates the circadian phase of clock gene expression. *Mol. Cell. Biol.* 28, 4080–4092.
- Nyfel, B., Michnick, S.W., and Hauri, H.P. (2005). Capturing protein interactions in the secretory pathway of living cells. *Proc. Natl. Acad. Sci. USA* 102, 6350–6355.
- Olsen, J.V., Vermeulen, M., Santamaria, A., Kumar, C., Miller, M.L., Jensen, L.J., Gnad, F., Cox, J., Jensen, T.S., Nigg, E.A., et al. (2010). Quantitative phosphoproteomics reveals widespread full phosphorylation site occupancy during mitosis. *Sci. Signal.* 3, ra3.
- Payne, K.J., and Dovat, S. (2011). Ikaros and tumor suppression in acute lymphoblastic leukemia. *Crit. Rev. Oncog.* 16, 3–12.

- Ramachandran, N., Raphael, J.V., Hainsworth, E., Demirkan, G., Fuentes, M.G., Rolfs, A., Hu, Y., and LaBaer, J. (2008). Next-generation high-density self-assembling functional protein arrays. *Nat. Methods* 5, 535–538.
- Richter, J., Schlesner, M., Hoffmann, S., Kreuz, M., Leich, E., Burkhardt, B., Rosolowski, M., Ammerpohl, O., Wagener, R., Bernhart, S.H., et al.; ICGC MMML-Seq Project (2012). Recurrent mutation of the *ID3* gene in Burkitt lymphoma identified by integrated genome, exome and transcriptome sequencing. *Nat. Genet.* 44, 1316–1320.
- Rozenblatt-Rosen, O., Deo, R.C., Padi, M., Adelmant, G., Calderwood, M.A., Rolland, T., Grace, M., Dricot, A., Askenazi, M., Tavares, M., et al. (2012). Interpreting cancer genomes using systematic host network perturbations by tumour virus proteins. *Nature* 487, 491–495.
- Rual, J.-F., Venkatesan, K., Hao, T., Hirozane-Kishikawa, T., Dricot, A., Li, N., Berriz, G.F., Gibbons, F.D., Dreze, M., Ayivi-Guedehoussou, N., et al. (2005). Towards a proteome-scale map of the human protein-protein interaction network. *Nature* 437, 1173–1178.
- Ruepp, A., Waegle, B., Lechner, M., Brauner, B., Dunger-Kaltenbach, I., Fobo, G., Frishman, G., Montrone, C., and Mewes, H.W. (2010). CORUM: the comprehensive resource of mammalian protein complexes—2009. *Nucleic Acids Res.* 38 (Database issue), D497–D501.
- Salwinski, L., Miller, C.S., Smith, A.J., Pettit, F.K., Bowie, J.U., and Eisenberg, D. (2004). The Database of Interacting Proteins: 2004 update. *Nucleic Acids Res.* 32 (Database issue), D449–D451.
- Stagljär, I., and Fields, S. (2002). Analysis of membrane protein interactions using yeast-based technologies. *Trends Biochem. Sci.* 27, 559–563.
- Stein, A., Céol, A., and Aloy, P. (2011). 3did: identification and classification of domain-based interactions of known three-dimensional structure. *Nucleic Acids Res.* 39 (Database issue), D718–D723.
- Stelzl, U., Worm, U., Lalowski, M., Haenig, C., Brembeck, F.H., Goehler, H., Stroedicke, M., Zenkner, M., Schoenherr, A., Koeppen, S., et al. (2005). A human protein-protein interaction network: a resource for annotating the proteome. *Cell* 122, 957–968.
- Stenson, P.D., Mort, M., Ball, E.V., Shaw, K., Phillips, A., and Cooper, D.N. (2014). The Human Gene Mutation Database: building a comprehensive mutation repository for clinical and molecular genetics, diagnostic testing and personalized genomic medicine. *Hum. Genet.* 133, 1–9.
- Stumpf, M.P., Thorne, T., de Silva, E., Stewart, R., An, H.J., Lappe, M., and Wiuf, C. (2008). Estimating the size of the human interactome. *Proc. Natl. Acad. Sci. USA* 105, 6959–6964.
- Thomas, G., Jacobs, K.B., Yeager, M., Kraft, P., Wacholder, S., Orr, N., Yu, K., Chatterjee, N., Welch, R., Hutchinson, A., et al. (2008). Multiple loci identified in a genome-wide association study of prostate cancer. *Nat. Genet.* 40, 310–315.
- Venkatesan, K., Rual, J.-F., Vazquez, A., Stelzl, U., Lemmens, I., Hirozane-Kishikawa, T., Hao, T., Zenkner, M., Xin, X., Goh, K.I., et al. (2009). An empirical framework for binary interactome mapping. *Nat. Methods* 6, 83–90.
- Vidal, M., Cusick, M.E., and Barabási, A.-L. (2011). Interactome networks and human disease. *Cell* 144, 986–998.
- Vogelstein, B., Papadopoulos, N., Velculescu, V.E., Zhou, S., Diaz, L.A., Jr., and Kinzler, K.W. (2013). Cancer genome landscapes. *Science* 339, 1546–1558.
- von Mering, C., Huynen, M., Jaeggi, D., Schmidt, S., Bork, P., and Snel, B. (2003). STRING: a database of predicted functional associations between proteins. *Nucleic Acids Res.* 31, 258–261.
- Wilhelm, M., Schlegel, J., Hahne, H., Moghaddas Gholami, A., Lieberenz, M., Savitski, M.M., Ziegler, E., Butzmann, L., Gessulat, S., Marx, H., et al. (2014). Mass-spectrometry-based draft of the human proteome. *Nature* 509, 582–587.
- Wölfel, T., Hauer, M., Schneider, J., Serrano, M., Wölfel, C., Klehmann-Hieb, E., De Plaen, E., Hankeln, T., Meyer zum Büschenfelde, K.H., and Beach, D. (1995). A p16INK4a-insensitive CDK4 mutant targeted by cytolytic T lymphocytes in a human melanoma. *Science* 269, 1281–1284.
- Woodsmith, J., and Stelzl, U. (2014). Studying post-translational modifications with protein interaction networks. *Curr. Opin. Struct. Biol.* 24, 34–44.
- Yang, X., Boehm, J.S., Yang, X., Salehi-Ashtiani, K., Hao, T., Shen, Y., Lubonja, R., Thomas, S.R., Alkan, O., Bhimdi, T., et al. (2011). A public genome-scale lentiviral expression library of human ORFs. *Nat. Methods* 8, 659–661.
- Yu, H., Tardivo, L., Tam, S., Weiner, E., Gebreab, F., Fan, C., Svrzikapa, N., Hirozane-Kishikawa, T., Rietman, E., Yang, X., et al. (2011). Next-generation sequencing to generate interactome datasets. *Nat. Methods* 8, 478–480.
- Zhang, Q.C., Petrey, D., Deng, L., Qiang, L., Shi, Y., Thu, C.A., Bisikirska, B., Lefebvre, C., Accili, D., Hunter, T., et al. (2012). Structure-based prediction of protein-protein interactions on a genome-wide scale. *Nature* 490, 556–560.
- Zhong, Q., Simonis, N., Li, Q.R., Charlotiaux, B., Heuze, F., Klitgord, N., Tam, S., Yu, H., Venkatesan, K., Mou, D., et al. (2009). Edgetic perturbation models of human inherited disorders. *Mol. Syst. Biol.* 5, 321.

Molecular Architecture of the 40S•eIF1•eIF3 Translation Initiation Complex

Jan P. Erzberger,* Florian Stengel, Riccardo Pellarin, Suyang Zhang, Tanja Schaefer, Christopher H.S. Aylett, Peter Cimermančić, Daniel Boehringer, Andrej Sali, Ruedi Aebersold, and Nenad Ban*

*Correspondence: jan.erzberger@mol.biol.ethz.ch (J.P.E.), ban@mol.biol.ethz.ch (N.B.)

<http://dx.doi.org/10.1016/j.cell.2014.11.001>

(Cell 158, 1123–1135; August 28, 2014)

In preparing the article above, we inadvertently assigned the coordinates for eIF3f to its homolog eIF3h and vice versa within the mammalian eIF3 model during PDB database deposition. This error led to improper positioning of these subunits in Figures 3E, 7F, and S2B. We present below revised panels for Figures 3 and 7 with the correct assignment for eIF3f and eIF3h. These figures, along with Figure S2, have been corrected online. The deposited PDB files have also been corrected and updated. In addition, we have added the missing legend for Figure S2B online. We apologize for any confusion that these errors may have caused.

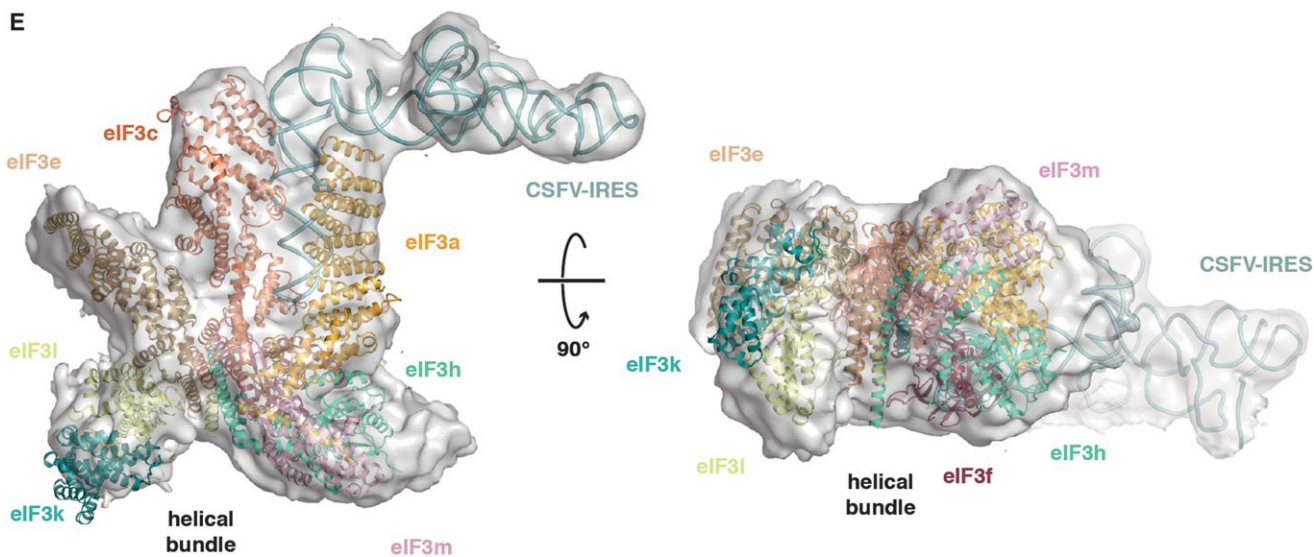


Figure 3. Docking of eIF3a/eIF3c in the PCI•MPN Core Density of Mammalian 43S and 43S•IRES EM Maps

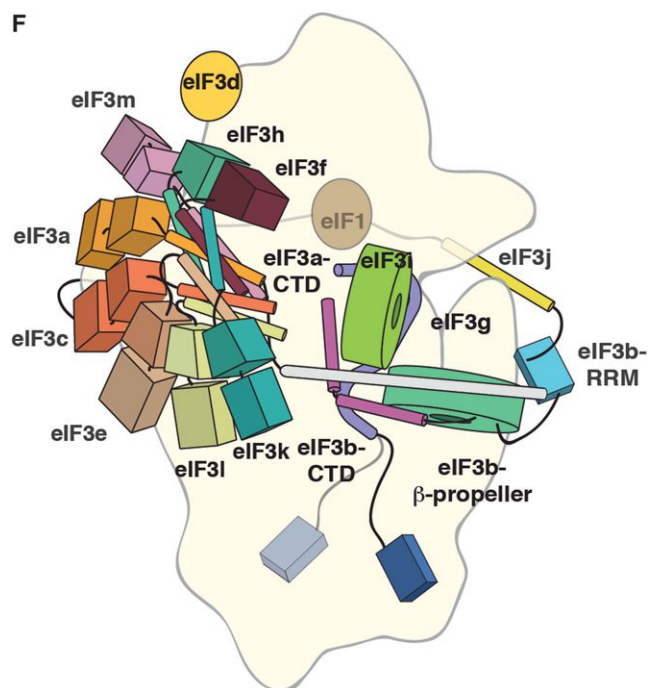


Figure 7. Placement and Interactions of eIF3 Components on 40S

ATRX Directs Binding of PRC2 to Xist RNA and Polycomb Targets

Kavitha Sarma, Catherine Cifuentes-Rojas, Ayla Ergun, Amanda del Rosario, Yesu Jeon, Forest White, Ruslan Sadreyev, and Jeannie T. Lee*

*Correspondence: lee@molbio.mgh.harvard.edu

<http://dx.doi.org/10.1016/j.cell.2014.11.010>

(Cell 159, 869–883; November 6, 2014)

During the final revision process, we inadvertently removed a key reference to prior work linking ATRX to macroH2A. On page 871, the sentence “ATRX caught our attention, as it was shown to be enriched on the Xi by immunofluorescence (Baumann and De La Fuente, 2009) and has an ATPase and helicase domain (Clynes et al., 2013; Ratnakumar and Bernstein, 2013)” should have cited the paper by Ratnakumar et al. (2012). The text and reference list have been corrected online.

REFERENCES

Ratnakumar, K., Duarte, L.F., LeRoy, G., Hasson, D., Smeets, D., Vardabasso, C., Bönisch, C., Zeng, T., Xiang, B., Zhang, D.Y., et al. (2012). ATRX-mediated chromatin association of histone variant macroH2A1 regulates α -globin expression. *Genes Dev.* 26, 433–438.



Where **Science** Meets the **City**

October 6-10 • Baltimore, MD

Invited Proposals and Workshops Accepted Through December 5, 2014

Seeking **invited session** proposals in:

- Emerging Genomic and Bioinformatic Methods and Tools
- Functional or Mechanistic Studies of Genetic Disease
- Comparative Genomics (Population or Species)
- Advances in Mendelian or Complex Trait Analysis
- WES or WGS in Clinical or Research Settings
- Evolutionary Genetics and Genomics
- Engagement of Stakeholders in Clinical Genetics/Genomics
- Social Issues, Education, and Policy
- and more!

June 11, 2015

**Abstract Submission
Deadline**



TRAINEES

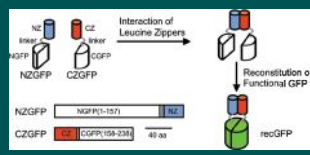
At least one slot has been set aside for an invited session that is proposed, moderated, and presented by trainees! When submitting a proposal, please mark the appropriate checkbox to indicate your status as a trainee.

ashg.org/2015meeting

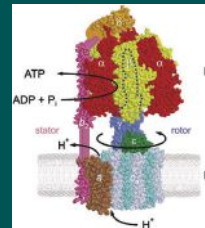
Cell Line: 2004–2014

MicroRNA biology
Bartel
Vol. 116, 281–297

First Resource



First Theory article



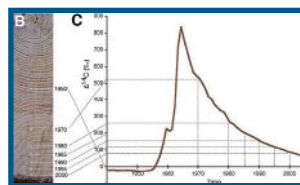
Cell gets a new look: Leading Edge is born



Microbiota maintains host homeostasis

Medzhitov
Vol. 118, 229–241 **AC**

"This paper proposed the notion that immune ignorance to the microbiota, the prevailing theory at the time, may not hold true in all cases."



Carbon dating the brain

Frisén
Vol. 122, 133–143

Four factors reprogram mouse fibroblasts

Takahashi and Yamanaka
Vol. 126, 663–676 **AC**

"The finding that iPSCs could also be derived from adult tail tip fibroblasts nailed this story."



2004

2005

2006

Open archive

NF-κB links inflammation and cancer

Karin
Vol. 118, 285–296

Discovery of MAVS

Chen
Vol. 122, 669–682

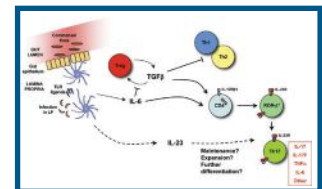
Histone demethylase

Shi
Vol. 119, 941–953 **AC**

"The inability to see global demethylation by LSD1... demonstrates the gene specific nature of demethylation...."

High-resolution view on DNA methylation

Jacobsen and Ecker
Vol. 126, 1189–1201



RORγt master of Th17 cells

Cua and Littman
Vol. 126, 1121–1133 **AC**

"...the claim that Th17 cells were a distinct population met with much skepticism.... [T]hat an orphan nuclear receptor RORγt was the transcription factor required for the differentiation of Th17 cells from that common progenitor resolved that controversy."

LEGEND

Landmark Paper
Essential Review
Cell Milestone
Publishing Innovation
AC Annotated Classic
 Nobel Prize

FORWARD FROM

40

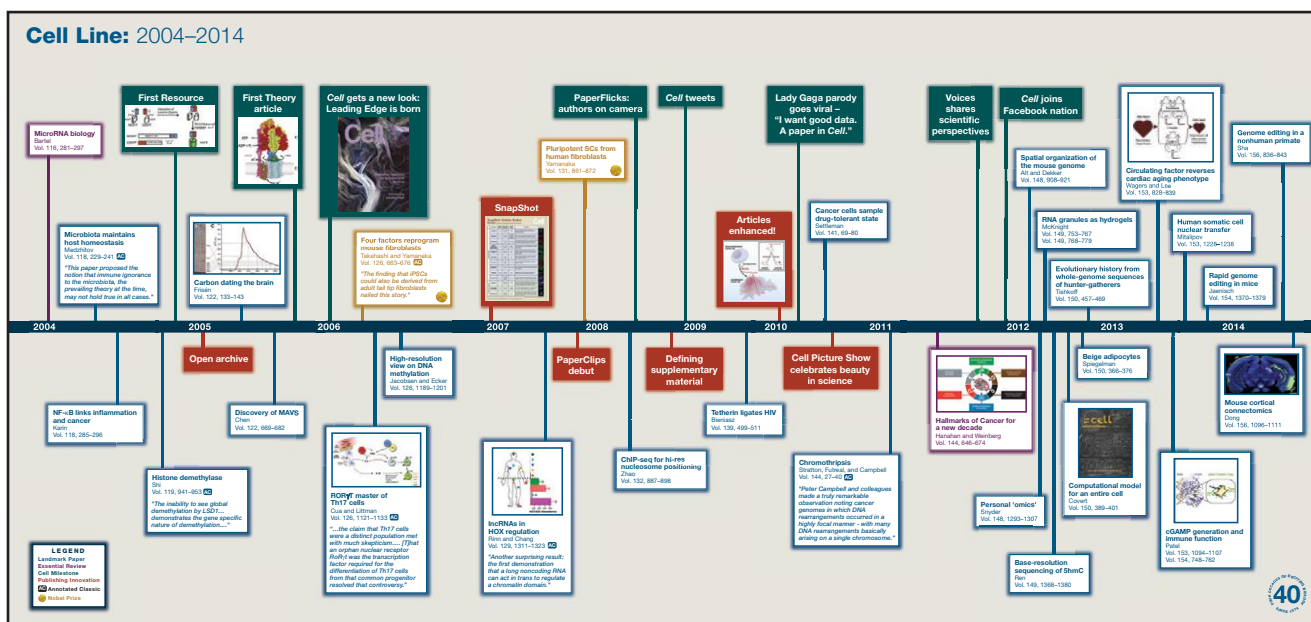


As we commemorate 40 years of publishing exciting biology we're also embracing the future. We will continue to lead the way with best-in-class editorial service, innovations in publishing, and new technology to address the evolving needs of scientists. Here's to the next 40 years!

www.cell.com/40/home

Cell

Cell Line: 2004–2014



2014 marks *Cell*'s 40th anniversary, and over the year we have looked back at how discoveries of the last four decades have molded our understanding of biology. The final decade of the Cell Line features a selection of the exceptional scientific work—both landmark papers and essential reviews. Select entries can be read as an “Annotated Classic,” which includes the original paper and accompanying reflections of a leading scientist, considering the work from our current vantage point. Our last installment includes a harbinger of the interplay between microbiota and mammalian hosts in 2004, revolutionary papers in 2006 and 2007 unlocking cellular reprogramming, the discovery of beige adipocytes in 2012, and the first example of CRISPR-based genome editing in a non-human primate in 2014. In addition to landmark publications, there were innovative developments at the journal in this decade, with the complete redesign of the print journal and the creation of Leading Edge in late 2005 and the restructuring of the online display of the article in 2010. Keeping pace with the changing nature of biological research, over the decade *Cell* added new article types, introduced guidelines for the organization of supplementary material and expanded the journal's web-based content to bring editors' and authors' excitement and perspective on individual papers to the readership.

An interactive version of the timeline, with links to the papers, full author lists, and Annotated Classics, is available at <http://dx.doi.org/10.1016/j.cell.2014.11.004>.

The *Cell* Editorial Team

PaperFlicks:
authors on camera

Cell tweets

Lady Gaga parody
goes viral –
“I want good data.
A paper in *Cell*.”

**Pluripotent SCs from
human fibroblasts**
Yamanaka
Vol. 131, 861–872

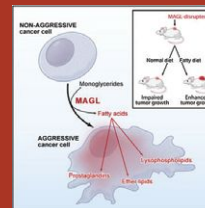


Snapshot

Snapshot: Cellular Bodies

Cell Type	Volume	Issue	Page Range	Lead Author	Journal
Embryonic stem cells	131	1	861–872	Shinya Yamanaka	Cell
Induced pluripotent stem cells	131	1	861–872	Shinya Yamanaka	Cell
Human embryonic stem cells	131	1	861–872	Shinya Yamanaka	Cell
Human embryonic stem cells	131	1	861–872	Shinya Yamanaka	Cell
Human embryonic stem cells	131	1	861–872	Shinya Yamanaka	Cell
Human embryonic stem cells	131	1	861–872	Shinya Yamanaka	Cell
Human embryonic stem cells	131	1	861–872	Shinya Yamanaka	Cell
Human embryonic stem cells	131	1	861–872	Shinya Yamanaka	Cell
Human embryonic stem cells	131	1	861–872	Shinya Yamanaka	Cell
Human embryonic stem cells	131	1	861–872	Shinya Yamanaka	Cell

Articles enhanced!



**Cancer cells sample
drug-tolerant state**
Settleman
Vol. 141, 69–80

2007

2008

2009

2010

2011

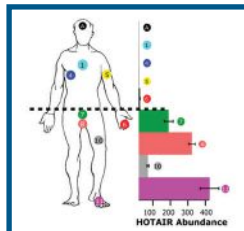
PaperClips
debut

**Defining
supplementary
material**

Cell Picture Show
celebrates beauty
in science

Tetherin ligates HIV
Bieniasz
Vol. 139, 499–511

**ChIP-seq for hi-res
nucleosome positioning**
Zhao
Vol. 132, 887–898



**lncRNAs in
HOX regulation**
Rinn and Chang
Vol. 129, 1311–1323 **AC**

*"Another surprising result:
the first demonstration
that a long noncoding RNA
can act in trans to regulate
a chromatin domain."*

Chromothripsis
Stratton, Futreal, and Campbell
Vol. 144, 27–40 **AC**
*"Peter Campbell and colleagues
made a truly remarkable
observation noting cancer
genomes in which DNA
rearrangements occurred in a
highly focal manner - with many
DNA rearrangements basically
arising on a single chromosome."*

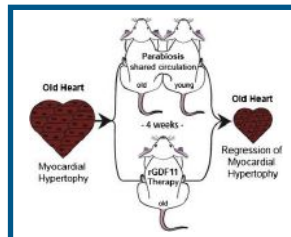
Voices
shares
scientific
perspectives

**Cell joins
Facebook nation**

**Spatial organization of
the mouse genome**
Alt and Dekker
Vol. 148, 908–921

RNA granules as hydrogels
McKnight
Vol. 149, 753–767
Vol. 149, 768–779

**Evolutionary history from
whole-genome sequences
of hunter-gatherers**
Tishkoff
Vol. 150, 457–469



**Circulating factor reverses
cardiac aging phenotype**
Wagers and Lee
Vol. 153, 828–839

**Genome editing in a
nonhuman primate**
Sha
Vol. 156, 836–843

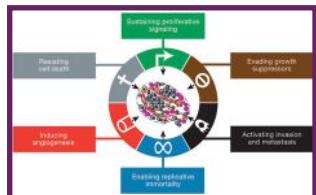
**Human somatic cell
nuclear transfer**
Mitalipov
Vol. 153, 1228–1238

**Rapid genome
editing in mice**
Jaenisch
Vol. 154, 1370–1379

2012

2013

2014

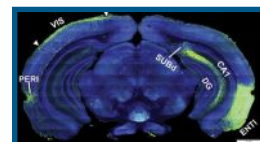


**Hallmarks of Cancer for
a new decade**
Hanahan and Weinberg
Vol. 144, 646–674

Beige adipocytes
Spiegelman
Vol. 150, 366–376



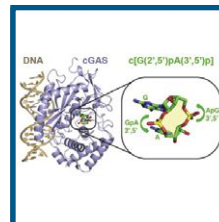
**Computational model
for an entire cell**
Covert
Vol. 150, 389–401



**Mouse cortical
connectomics**
Dong
Vol. 156, 1096–1111

Personal 'omics'
Snyder
Vol. 148, 1293–1307

**Base-resolution
sequencing of 5hmC**
Ren
Vol. 149, 1368–1380



**cGAMP generation and
immune function**
Patel
Vol. 153, 1094–1107
Vol. 154, 748–762

40
FOUR DECADES OF EXCITING BIOLOGY
SINCE 1974

Join or Renew Your Membership Today!



THE
AMERICAN
SOCIETY
OF HUMAN
GENETICS

New for 2013, become a Regular Member of ASHG and get a discount on Regular Membership with the European Society of Human Genetics!

\$25 - \$70 — Trainee Members

\$190 — Regular Members
multi-year discounts now available



Membership benefits include:

- Reduced fees for the 2014 ASHG Annual meeting
- Subscription to *The American Journal of Human Genetics*
- Influence on Public Policy Issues
- Access to Educational Materials
- Free access to invited session recordings from past ASHG Annual Meetings through the membership web portal
- Discount to Journal Subscriptions (*Nature*, *Nature Reviews Genetics*, and more!)

For more information visit:

www.ashg.org/join

If you have questions regarding membership status, please contact us:

The American Society of Human Genetics
9650 Rockville Pike • Bethesda, MD 20814-3998
1.866.HUMGENE • Membership: scelia@ashg.org



COLUMBIA UNIVERSITY
IN THE CITY OF NEW YORK

The 2015 Louisa Gross Horwitz Prize for Biology or Biochemistry

The Louisa Gross Horwitz Prize was established under the will of the late S. Gross Horwitz through a bequest to Columbia University and is named to honor the donor's mother. Louisa Gross Horwitz was the daughter of Dr. Samuel David Gross (1805-1889), a prominent surgeon of Philadelphia and author of the outstanding *Systems of Surgery* who served as President of the American Medical Association.

Each year since its inception in 1967, the Louisa Gross Horwitz Prize has been awarded by Columbia University for outstanding basic research in the fields of biology or biochemistry. The purpose of this award is to honor a scientific investigator or group of investigators whose contributions to knowledge in either of these fields are deemed worthy of special recognition.

The Prize consists of an honorarium and a citation which are awarded at a special presentation event. Unless otherwise recommended by the Prize Committee, the Prize is awarded annually. Dr. James P. Allison, University of Texas MD Anderson Cancer Center was the 2014 awardee.

QUALIFICATIONS FOR THE AWARD

The Prize Committee recognizes no geographical limitations. The Prize may be awarded to an individual or a group. When the Prize is awarded to a group, the honorarium will be divided among the recipients, but each member will receive a citation. Preference will be given to work done in the recent past.

Nominations must be submitted electronically at: <http://www.cumc.columbia.edu/research/horwitz-prize>

All communications and materials must be written in the English language.

Re-nomination(s) are by invitation only.

Self-nominations are not permitted.

Nominations should include:

- 1) A summary, no more than 500 words long, of the research on which this nomination is based.
- 2) A summary, no more than 500 words long, of the significance of this research in the fields of biology or biochemistry.
- 3) A brief biographical sketch of the nominee, including positions held and awards received by the nominee.
- 4) A listing of up to ten of the nominee's most significant publications relating to the research noted under item 1.
- 5) A copy of the nominee's curriculum vitae.

Deadline date: January 30, 2015



BIOMEDICAL DISCOVERY SCIENTIST

The Department of Biochemistry and Molecular Biology at Mayo Clinic in Rochester, MN is seeking an outstanding scientist to align with its newly designated Center for Biomedical Discovery. The individual will develop and/or maintain a nationally and internationally recognized extramurally funded research program. Candidates should hold a Ph.D., M.D., or M.D./Ph.D. and have expertise in fibrosis, senescence, or cancer.

Mayo Clinic has been recognized as the best hospital in the nation for 2014-2015 by *U.S. News and World Report* and is consistently recognized by *FORTUNE* magazine as one of the "100 Best Companies to Work For." We are an integrated, multidisciplinary academic medical center with comprehensive programs in medical education and research that span across three group practice sites. We support a vibrant research enterprise, with programs in clinical, basic and population sciences. In 2013, the institution received nearly \$370 million in extramural research awards, supplemented with over \$275 million of institutional support. Mayo Clinic has a large number of outstanding laboratory and shared resource facilities, including a Medical Genomics Facility, a Mouse Modeling Facility, Molecular and Microscopic Core Facilities, and unique Human Disease Tissue Banks. We offer a highly competitive compensation package with sustained intramural funding, capital equipment funding, technical and computational resources, and exceptional benefits.

To apply online, please visit: <http://mayoclinic.in/1zfshFU>

For further information please contact: Edward B. Leof, Ph.D.,
Search Committee Chair at Schilbe.Jennifer@mayo.edu

Heal the sick, advance the science, share the knowledge.

©2014 Mayo Foundation for Medical Education and Research. Post offer/pre-employment drug screening is required.
Mayo Clinic is an equal opportunity educator and employer (including veterans and persons with disabilities).



STANFORD UNIVERSITY DEPARTMENT OF CHEMICAL AND SYSTEMS BIOLOGY

The Department of Chemical and Systems Biology at Stanford University School of Medicine invites applications for a tenure-track position at the ASSISTANT PROFESSOR level. We are particularly interested in candidates with a strong interdisciplinary record in the broad areas of chemical biology, systems biology, and/or cellular and molecular biology in normal and disease states. Stanford offers an outstanding environment for creative interdisciplinary biomedical research. The main criterion for appointment in the University Tenure Line is a major commitment to research and teaching. For more information on our department, please visit: <http://chemsysbio.stanford.edu/>.

Candidates should have a Ph.D. and/or M.D. degree and postdoctoral research experience. Applications should include a cover letter addressed to

Tobias Meyer, Ph.D., Search Committee Chair; a curriculum vitae, publication list, description of future research plans, and at least 3 letters of reference. All materials should be submitted online to Academic Jobs Online at <https://academicjobsonline.org/ajob/jobs/4602>.

To ensure full consideration, please submit your applications by December 1, 2014. Late applications may be considered.

Questions should be addressed to Kathy Johnson,
Faculty Affairs Administrator at csbsearch@stanford.edu.

Stanford University in an equal opportunity employer and is committed to increasing the diversity of its faculty. It welcomes nominations of and applications from women, members of minority groups, protected veterans and individuals with disabilities, as well as from others who would bring additional dimensions to the university's research, teaching and clinical missions

Be their next big discovery.

careers.cell.com

Where top life sciences talent is searching
for their next opportunity

Cell Career Network

Discover the best jobs in life science

To showcase your job openings on Cell Career Network,
contact: advertising@cell.com

Postdoctoral Program

Novartis Institutes for
BioMedical Research

Do you have a passion for innovative
fundamental research in drug discovery?

We are seeking creative postdoctoral scholars to
conduct exciting research in biology, chemistry,
and computational sciences at the frontier of
drug discovery.

Postdocs at NIBR who will subsequently pursue
faculty positions in academia are eligible for
start-up funds through our new and highly
competitive Young Investigator Awards.

Visit <http://postdoc.nibr.com> to view our
mentors' research profiles and to apply.

Assistant Professor

Departments of Cancer Biology, Dana-Farber Cancer Institute
and Cell Biology, Harvard Medical School

The Departments of Cancer Biology at the Dana-Farber Cancer Institute and Cell Biology at Harvard Medical School invite applicants for a tenure-track faculty position at the rank of Assistant Professor or Associate Professor. We are seeking individuals with a demonstrated potential for imaginative research and who propose to work on exciting problems in any area of Molecular Metabolism. We are especially interested in candidates who utilize metabolomics, proteomics or both to uncover novel metabolic pathways in health and diseased states. The successful candidate will be expected to direct innovative and independent research and participate in the teaching of graduate and/or medical students. Our highly interactive environment provides the opportunity to engage and collaborate with other dedicated researchers both within the Division of Metabolism and Chronic Disease of the Cancer Biology Department (chaired by Bruce Spiegelman) and throughout the diverse Harvard research community. Significant scholarly and scientific resources will be made available for this appointment. Applicants will be housed in new space at the Dana-Farber Cancer Institute. For further information about our Department, please see our web page: <http://www.dana-farber.org/Research/Departments-and-Centers/Department-of-Cancer-Biology.aspx>

Applicants should submit electronic copies of their curriculum vitae, a description of research accomplishments and future research interests (three pages maximum), and ask at least three references to provide letters of recommendation. These materials should be submitted using the following link: <https://academicpositions.harvard.edu/postings/5815>. Please contact Kim Wilkinson (Kim_Wilkinson@dfci.harvard.edu) with any questions regarding submission of documents.

Applications must be received by: February 28, 2015.

We are an equal opportunity employer and all qualified applicants will receive consideration for employment without regard to race, color, religion, sex, national origin, disability status, protected veteran status, or any other characteristic protected by law.



HARVARD
MEDICAL SCHOOL



USC University of
Southern California

Faculty Position

Division of Hematology, Department of Medicine
University of Southern California Keck School of Medicine

Supported by a \$58 million endowment, the Jane Anne Nohl Division of Hematology and Center for the Study of Blood Diseases and the USC Norris Comprehensive Cancer Center at the Keck School of Medicine seek faculty members interested in basic, translational and clinical research in novel treatment approaches for blood cancers. Candidates working in the areas of chimeric antigen receptors, allogeneic stem cell graft engineering, adoptive cell therapy, TCR gene therapy, immune mobilizing monoclonal T cells against cancer (ImmTac), peptide vaccination, antibody drug conjugates, bispecific antibodies, molecularly targeted drugs or other novel treatment approaches against blood cancers are strongly encouraged to apply. Preference will be given to applicants who have MD or MD/PhD degrees, BC/BE in Hematology and/or Medical Oncology and eligibility for a California medical license. Successful candidates will receive competitive compensation and appointment at the appropriate academic level (Assistant/Associate/Full Professor) commensurate with experience and accomplishments.

Applicants should send letter of interest, CV and names of three references to
Preet Chaudhary, MD, PhD
Chief of Nohl Division of Hematology
Keck School of Medicine of USC
1441 Eastlake Avenue, NOR 3470, MC 9172
Los Angeles, CA 90033
Email: Preet.Chaudhary@med.usc.edu.

Assistant/Associate Professor of Microbiology and Immunobiology Metabolism/Microbiome

The Joslin Diabetes Center, Boston, Massachusetts, invites applications from laboratory-based investigators with a focus on understanding reciprocal interactions between the microbiome and metabolism. The successful applicant will be recommended for full-time appointment as a tenure-track Assistant or Associate Professor, commensurate with accomplishments. This individual will develop an independent research program aimed at mechanistic investigations that will clarify the influence of microbiota on metabolic disorders and the development of diabetes and its complications, collaborating closely with faculty at the Joslin and Harvard Medical School interested in obesity, autoimmunity and inflammation. The candidate will be based at the Joslin Diabetes Center and will be a member of the Department of Microbiology and Immunobiology, Harvard Medical School.

This position includes a highly competitive compensation and laboratory start-up package. Applicants must have an MD and/or PhD and a proven track record of outstanding, innovative research.

Interested candidates should submit curriculum vitae, reprints of up to three of their most significant publications, a brief statement of present and future research plans (3-5 pages), and three letters of reference to **Professor Amy Wagers, Chair of the Search Committee**.

Materials should be submitted via email to:
microbiome@joslin.harvard.edu.

The submission deadline for receipt of all application materials is December 15, 2014. We strongly encourage applications from women and minority candidates. Joslin is the world's largest diabetes research and clinical care organization, and an independent, nonprofit institution affiliated with Harvard Medical School. Joslin is dedicated to ensuring that people with diabetes live long, healthy lives and offers real hope and progress toward diabetes prevention and a cure.

Our mission is to prevent, treat and cure diabetes.
Our vision is a world free of diabetes and its complications.

Joslin Diabetes Center is an equal opportunity employer and all qualified applicants will receive consideration for employment with regard to race, color, religion, sex, national origin, disability status, protected veteran status, or any other characteristic protected by law.



Joslin Diabetes Center

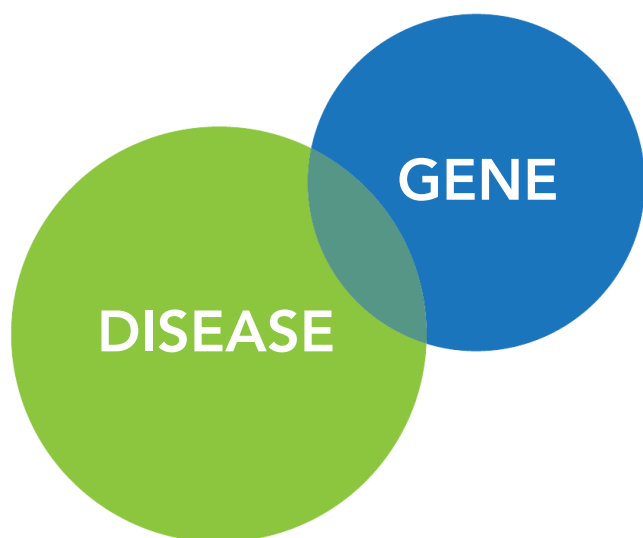


HARVARD
MEDICAL SCHOOL

UNCOVER NEW INSIGHTS TO HELP YOU IDENTIFY POTENTIAL BIOLOGICAL TARGETS

Target Insights is a support tool that helps users find important facts that other search tools may miss, because it is based upon an automated analysis of the full text of scientific articles. Target Insights gets you to the right target quicker and with more confidence than other tools.

Developed by Elsevier to meet increasing demands on speed and efficiency it enables target identification and validation teams to discover new insights and update their understanding of the disease model by revealing crucial information which is normally obscured and hard-to-find, in three steps... [Identify](#), [Prioritize](#), [Validate](#).



For the latest information visit
targetinsights.com/info

Copyright © 2012 Elsevier Properties SA. All rights reserved.
Target Insights™ is owned by Elsevier Properties SA and used under license.

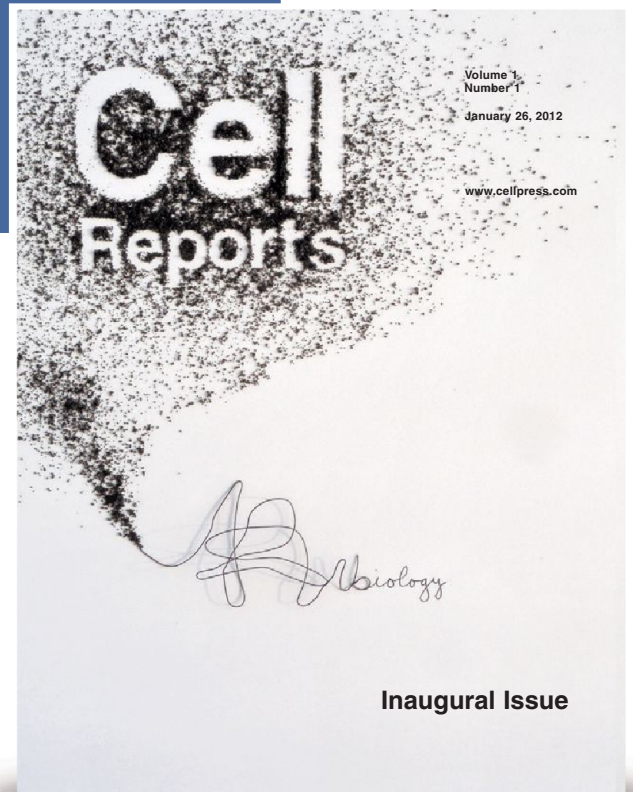


Discover high-quality, open-access research

Cell Reports features:

- High-quality, cutting-edge research
- A focus on short, single-point papers called Reports
- Broad scope covering all of biology
- Flexible open-access policy
- A highly engaged editorial board
- A distinguished advisory board
- New papers online weekly

cellreports.cell.com



Cell Press content is widely accessible

At Cell Press we place a high priority on ensuring that all of our journal content is widely accessible and on working with the community to develop the best ways to achieve that goal.

Here are just some of those initiatives...

Open archives

We provide free access to Cell Press research journals 12 months following publication

Access for developing nations

We provide free & low-cost access through programs like Research4Life

Open access journal

We launched Cell Reports - a new Open Access journal spanning the life sciences

Funding body agreements

We work cooperatively and successfully with major funding bodies

Public access

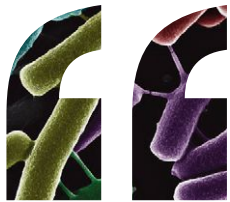
Full-text online via ScienceDirect is also available to the public via walk in user access from any participating library

Submission to PubMed Central

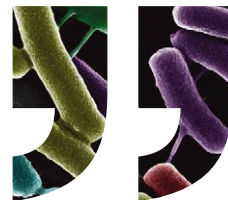
Cell Press deposits accepted manuscripts on our authors' behalf for a variety of funding bodies, including NIH and HHMI, to PubMed Central (PMC)

www.cell.com/cellpress/access





We recognize what matters in science.



Our editors select reviewers appropriate for your paper and will help you interpret their comments – striking a balance between the ideal and the reasonable.

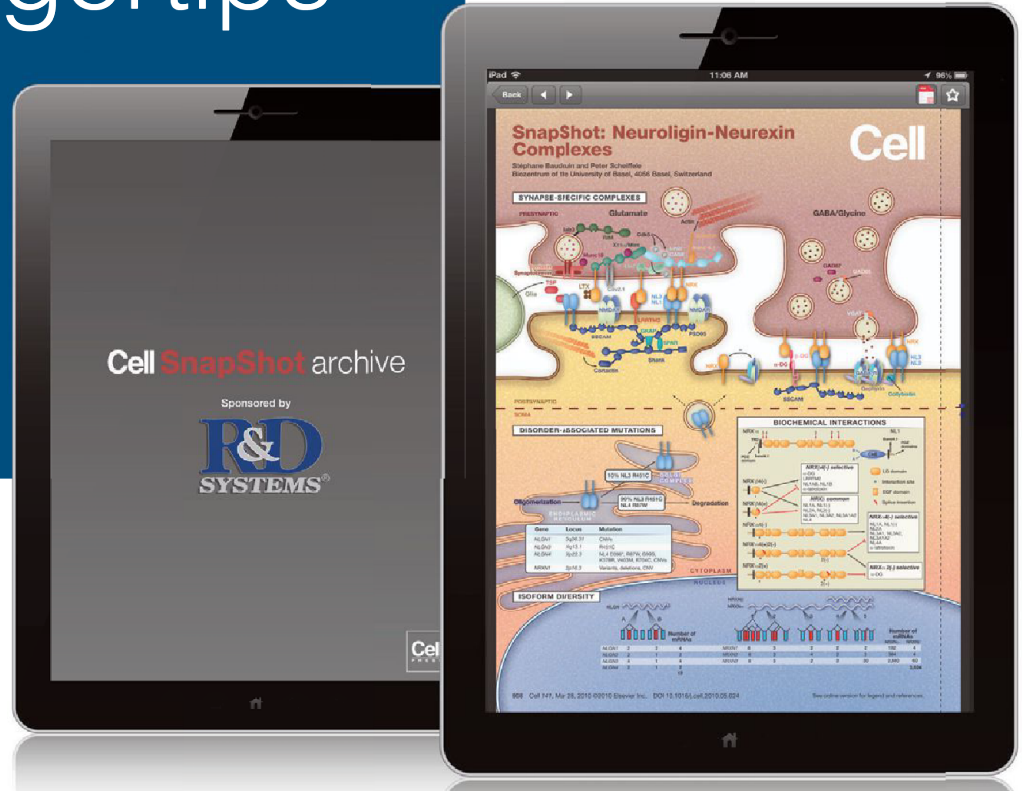
When you choose Cell Press you get the attention you deserve.

For more information visit: www.cell.com/values

Your **work** is our **life**



Cell Press SnapShots at your fingertips



Access our Snapshot collection as well as new Snapshots every two weeks for free! SnapShots are handy, one-page reference guides that include up-to-date tables of nomenclature and glossaries, signaling pathways, schematic diagrams of cellular processes, concepts in development and disease, and much more.

- All free
- Beautiful images
- Great for studying
- Searchable



DOWNLOAD
THE APP TODAY AT
cell.com/mobile

SPONSORED BY
R&D
SYSTEMS®

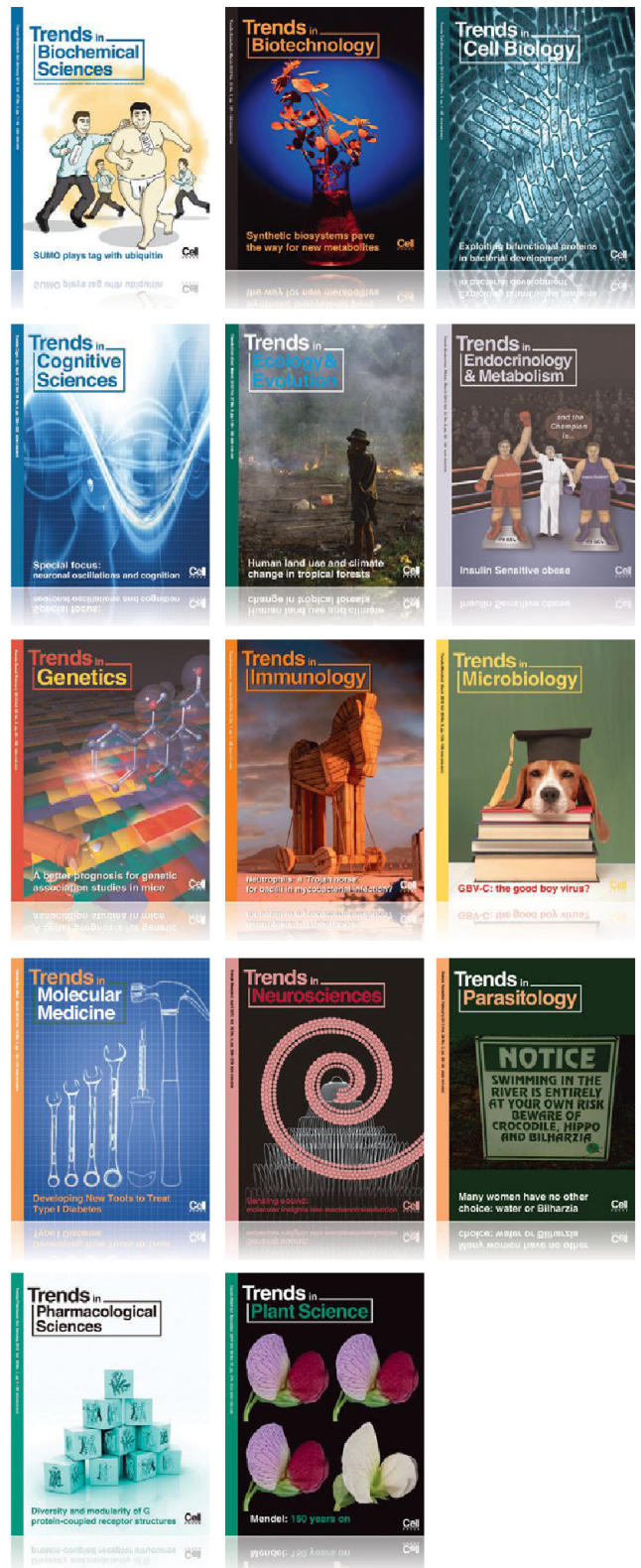
Cell
PRESS

Discover the latest Trends in your field

Cell Press Trends journals feature:

- Cutting-edge Review and Opinion articles
- Authoritative, succinct and accessible content
- Discussion, analysis and debate

For more information visit
cell.com/trends



Trends



Look Again...



Discover More.

- Access to the 14 Cell Press primary research journals and 14 Trends reviews titles, all on the same platform
- Improved, more robust article and author search
- Video, animations and sound files
- Easy to navigate home page, articles pages and archive

Cell
P R E S S

Trends

www.cell.com

Supported by:



Even more ways to access Cell Press journals



App Features:

- Available to subscribers at no additional charge
- Free 30-day trial to all 30 Cell Press journals
- Interactive reading experience
- Browse abstracts and full-text articles
- Download now, read later
- Make article notes
- Share with your friends and colleagues
- Now on the iPhone, iPad, Android tablet and Android phone



Download the app at
cell.com/mobile

CellPress



Simply stunning EVOS[®] cell imaging systems

Minimize the complexities of microscopy without compromising performance
Smarter systems | Easier cell imaging | Faster results

Compare systems at lifetechnologies.com/simplystunning

life
technologies

A Thermo Fisher Scientific Brand

For Research Use Only. Not for use in diagnostic procedures. © 2014 Thermo Fisher Scientific Inc. All rights reserved.
All trademarks are the property of Thermo Fisher Scientific and its subsidiaries unless otherwise specified. C0118976 0814